BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

#13 Attachment
09/729653

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/53, 15/57, 9/02, 9/64, C12Q 1/32, 1/37, C07K 16/40, A61K 38/44, 38/48, 48/00

A2

(11) International Publication Number: WO 00/65067

(43) International Publication Date: 2 November 2000 (02.11.00)

(21) International Application Number: PCT/US00/10920

(22) International Filing Date: 21 April 2000 (21.04.00)

(30) Priority Data:

60/130,778	23 April 1999 (23.04.99)	US
60/151,585	30 August 1999 (30.08.99)	US
60/174,003	30 December 1999 (30.12.99)	US
60/177,751	24 January 2000 (24.01.00)	US

(71) Applicant *(for all designated States except US)*: UNIVERSITY OF WASHINGTON [US/US]; Office of Technology Transfer, 1107 N.E. 45th Street, Suite 200, Seattle, WA 98105–4631 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: NELSON, Peter, S. [US/US]; 5121 27th Avenue N.E., Seattle, WA 98105 (US). HOOD, Leroy [US/US]; 6411 N.E. Windermer Drive, Seattle, WA 98105 (US). LIN, Biaoyang [CN/US]; 13716 15th Avenue N.E., #204, Seattle, WA 98125 (US).

(74) Agent: BRADLEY, R., Douglas; Christensen O'Connor Johnson & Kindness PLLC, Suite 2800, 1420 Fifth Avenue, Seattle, WA 98101 (US).

(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: PROSTATE-SPECIFIC POLYNUCLEOTIDES, POLYPEPTIDES AND THEIR METHODS OF USE

(57) Abstract

The invention provides isolated polynucleotides encoding prostate-specific, androgen-regulated polypeptides. The invention also provides substantially pure polynucleotides corresponding to genomic regulator regions of prostate-specific, androgen-regulated polynucleotides. Fragments and probes of polynucleotides thereof are also provided. The invention further provides a method of diagnosing or predicting the susceptibility of a prostate neoplastic condition in an individual suspected of having a neoplastic condition of the prostate. The method consists of: (a) obtaining a fluid or prostate sample of the individual; (b) determining the expression level of the prostate-specific, androgen-regulated polynucleotide or polypeptide, and (c) comparing the expression levels of the prostate-specific, androgen-regulated polynucleotide or polypeptide to expression levels from a normal fluid sample, from normal prostate cells or from an androgen-dependent cell line, wherein a two-fold change in expression level of the prostate-specific, androgen-regulated polynucleotide or polypeptide in the individual fluid or prostate sample as compared to the normal fluid or normal prostate cells or an androgen-dependent cell line indicates the presence of a prostate neoplastic condition. Methods of identifying compounds that selectively inhibit or increase prostate-specific polypeptides of the invention and a method of treating or reducing the progression of a prostate neoplastic condition are also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

PROSTATE-SPECIFIC POLYNUCLEOTIDES, POLYPEPTIDES AND THEIR METHODS OF USE

This invention was made with government support under grant number KO8 CA75173-01A1 awarded by the National Institutes of Health. The United States Government has certain rights in this invention.

Related Applications

This application claims the benefit of U.S. provisional patent application Serial No. 60/130,778, filed on April 23, 1999, Serial No. 60/151,585, filed on August 30, 1999, Serial No. 60/174,003, filed on December 30, 1999, and Serial No. 60/177,751, filed on January 24, 2000.

Field of the Invention

This invention relates generally to prostate cancer and, more specifically, to androgen-regulated, prostate-specific, nucleic acid molecules, proteins and antibodies that can be used to diagnose and treat prostate cancer.

Background of the Invention

Cancer is currently the second leading cause of mortality in the United States. However, it is estimated that by the year 2000 cancer will surpass heart disease and become the leading cause of death in the United States. Prostate cancer is the most common non-cutaneous cancer in the United States and the second leading cause of male cancer mortality.

Cancerous tumors result when a cell escapes from its normal growth regulatory mechanisms and proliferates in an uncontrolled fashion. As a result of such uncontrolled proliferation cancerous tumors usually invade neighboring tissues and spread by lymph or blood stream to create secondary or metastatic growths in other tissues. If untreated, cancerous tumors follow a fatal course. Prostate cancer, due to its slow growth profile, is an excellent candidate for early detection and therapeutic intervention.

During the last decade, most advances in prostate cancer research have focused on prostate specific antigen (PSA), a member of the serine protease family that exhibits a prostate-specific expression profile. Serum PSA remains the most widely used tumor marker for monitoring prostate cancer, but its specificity is limited by a high frequency of falsely elevated values in men with benign prostatic hyperplasia (BPH). Other biomarkers of prostate cancer progression have proven to be of limited clinical use in recent surveys because they are not uniformly elevated in men with advanced prostate cancer. Due to the limitations of currently available biomarkers, the identification and characterization of prostate specific genes is essential to the development of more accurate diagnostic methods and therapeutic targets. The clinical potential of novel tumor markers can be optimized by either utilizing them in combination with other tumor markers or by themselves in the development of diagnostic and treatment modalities.

Androgens are a class of C19 steroids that are essential for the development, growth, and maintenance of the prostate. Androgens exert their effects on the prostate target cells via the intracellular androgen receptor (AR). The AR facilitates androgen-induced regulation of genes involved in cellular proliferation and differentiation. As is the case with normal prostate development, primary prostatic cancers are largely dependent on androgens for growth and survival. Imbalance in androgen synthesis and degradation in prostate cells can lead to excess androgen, causing excessive cell growth as seen in benign prostate hyperplasia (BPH) and prostate cancer. Prostate-specific genes that contain androgen receptor elements (AREs) necessary for androgen induction include PSA, which contains two AREs, and human prostate-specific kallikrein (hKLK2). Despite clinical evidence that control of proper intracellular androgen levels in prostate cells is critical to a healthy prostate, the molecular components underlying the development and progression of prostate cancer are poorly understood. Identification of the components controlling androgen-regulation of the prostate is important for the development of new treatment modalities to cure prostate neoplastic conditions.

Thus, there exists a need for identification of additional genes involved in androgen-regulation of the prostate. In addition, there exists a need for identification of additional prostate specific genes that can be used as diagnostic markers and therapeutic targets for prostate cancer. The present invention satisfies this need and provides related advantages as well.

Summary of the Invention

In accordance with the foregoing, cDNA molecules that are predominantly expressed in the prostate gland have been isolated and sequenced, and the corresponding amino acid sequences have been deduced. Accordingly, the present invention relates to isolated, recombinant polypeptides that are expressed in the prostate gland, and to isolated polynucleotide sequences which are predominantly expressed in the prostate

gland, such as the sequences designated: SEQ ID NO:1, which encodes ARSDR1, a short chain dehydrogenase/reductase having the amino acid sequence SEQ ID NO:2; polynucleotide SEQ ID NO:3, which encodes TMPRSS2, a serine protease having the amino acid sequence SEQ ID NO:4; polynucleotide SEQ ID NO:5, which encodes PART-1, a polypeptide of unknown function having the amino acid sequence SEQ ID NO:6; and polynucleotide SEQ ID NO:7, which encodes 8C3, a polypeptide of unknown function.

In one aspect, the present invention provides an isolated polynucleotide capable of hybridizing under stringent conditions to at least 15 contiguous nucleotides of a polynucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10 and SEQ ID NO:11.

In another aspect, the present invention provides a substantially pure polynucleotide probe comprising at least 15 contiguous nucleotides of a polynucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10 and SEQ ID NO:11, or a fragment thereof.

In yet another aspect, the present invention provides a substantially pure polypeptide comprising substantially an amino acid sequence selected from the group consisting of the sequences shown as SEQ ID NO:2, SEQ ID NO:6, and functional fragments thereof.

Another embodiment of the invention provides an antibody that specifically binds to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, and SEQ ID NO:6, or a fragment thereof.

The invention further provides a method of diagnosing or predicting the susceptibility of a prostate neoplastic condition in an individual suspected of having a neoplastic condition of the prostate. The method is performed by:

(a) obtaining a fluid sample from an individual;

(b) determining an expression level of at least one polypeptide selected from the group consisting of ARSDR1, TMPRSS2, and PART-1; and

(c) comparing said determined expression level of said chosen polypeptide to a normal expression level of said chosen polypeptide from a normal fluid sample, wherein said measured expression level for said chosen polypeptide of 2-fold or more from said fluid sample from said individual compared to said normal expression level indicates the presence of a prostate neoplastic condition. Alternatively, the method can be performed by obtaining a prostate cell sample from the individual, determining an expression level of one of the inventive polypeptides in the prostate cell sample, and comparing the prostate expression level to a normal expression level of the

corresponding inventive polypeptide from normal prostate cells or from an androgen-dependent cell line. Again, a 2-fold or more increase in expression of the inventive polypeptide from the prostate cell sample from the individual compared to the normal expression level indicates the presence of a prostate neoplastic condition.

Methods of identifying compounds that inhibit or increase the activity of the inventive polypeptides and a method of treating or reducing the progression of a prostate neoplastic condition are also provided.

Detailed Description of the Preferred Embodiment

This invention is directed to prostate localized polypeptides and encoding polynucleotide molecules. Promoter and regulatory regions of the prostate expressed transcripts are also included. More specifically, four different androgen-responsive polynucleotides and polypeptides are provided: a polynucleotide having the nucleotide sequence shown in SEQ ID NO:1, that encodes ARSDR1, a short-chain dehydrogenase/reductase 1 having the polypeptide sequence of SEQ ID NO:2; a polynucleotide having the nucleotide sequence shown in SEQ ID NO:3 that encodes TMPRSS2, a prostate-specific serine protease having the amino acid sequence presented in SEQ ID NO:4, and two polynucleotides having the nucleic acid sequences represented in SEQ ID NOS:5 and 7, respectively, that encode polypeptides of unknown function. Polynucleotide SEQ ID NOS:5 encodes a polypeptide having the amino acid sequences shown in SEQ ID NOS:6. The polypeptides encoded by the androgen-responsive polynucleotides of the present invention are useful as both diagnostic markers for neoplastic conditions of the prostate and as targets for therapy. Polynucleotides corresponding to the expressed transcripts or promoters and regulatory regions are similarly applicable in both diagnostic and therapeutic procedures.

In one embodiment, the invention is directed to polynucleotide transcripts of an androgen regulated polynucleotide encoded by one of the nucleotide sequences shown in SEQ ID NOS:1, 3, 5, and 7. The invention also pertains to 5' promoter and regulatory regions shown in SEQ ID NO:8 (nucleotides 1 to 3,113), SEQ ID NO:9, SEQ ID NO:11 and a 3' untranslated region (UTR) of TMPRSS2 (SEQ ID NO:10). The inventive polynucleotides, fragments of the polynucleotides and short oligonucleotides corresponding to unique sequences are useful in a variety of diagnostic procedures which employ probe hybridization methods. One advantage of employing nucleic acid hybridization in diagnostic procedures is that very little sample can be used because the analyte nucleic acid can be amplified to many copies by, for example, polymerase chain reaction (PCR) or other well known methods in the art for polynucleotide amplification and synthesis.

In another embodiment, the invention is directed to substantially pure polypeptides and functional fragments thereof that are encoded by the polynucleotides of

the invention. In particular, the inventive polypeptides can be used to prepare antibodies. ARSDR1, TMPRSS2, and PART-1 specific antibodies can be used, following a variety of methods that are well known in the art, to diagnose prostate cancer.

In another embodiment, the invention is directed to methods for diagnosing prostate neoplastic conditions. The short-chain dehydrogenase/reductase of the invention is primarily expressed in prostate cells and becomes elevated in response to androgens. As such, the polynucleotide sequences of the present invention are applicable alone or in combination with other molecules, as a specific marker for prostate cells and prostate neoplastic conditions.

As used herein, the term "nucleotide" means a monomeric unit of DNA or RNA containing a sugar moiety (pentose), a phosphate and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of pentose) and that combination of base and sugar is called a nucleoside. The base characterizes the nucleotide with the four bases of DNA being adenine ("A"), guanine ("G"), cytosine ("C") and thymine ("T"). Inosine ("I") is a synthetic base that can be used to substitute for any of the four, naturally-occurring bases (A, C, G or T). The four RNA bases are A,G,C and uracil ("U"). The nucleotide sequences described herein comprise a linear array of nucleotides connected by phosphodiester bonds between the 3' and 5' carbons of adjacent pentoses.

"Oligonucleotide" refers to short length single or double stranded sequences of deoxyribonucleotides linked via phosphodiester bonds. The oligonucleotides are chemically synthesized by known methods and purified, for example, on polyacrylamide gels.

The term "hybridize under stringent conditions", and grammatical equivalents thereof, means that a polynucleotide molecule that has hybridized to a target polynucleotide molecule immobilized on a DNA or RNA blot (such as a Southern blot or Northern blot) remains hybridized to the immobilized target molecule on the blot during washing of the blot under stringent conditions. In this context, exemplary hybridization conditions are: hybridization at 65°C in 5.0 X SSC, 1% sodium dodecyl sulfate, for 16 hours (lower stringency hybridizations preferably utilize 6.0 X SSC, 1% sodium dodecyl sulfate, at 20°C to 30°C for 16 hours). Exemplary very high stringency conditions for washing DNA or RNA blots are: two washes of fifteen minutes each at 20°C to 30°C in 2.0 X SSC, followed by two washes of twenty minutes each at 65°C in 0.5 X SSC. Exemplary high stringency conditions for washing DNA or RNA blots are: two washes of twenty minutes each at 20°C to 30°C in 2.0 X SSC, followed by one wash of thirty minutes at 55°C in 1.0 X SSC. Exemplary moderate stringency conditions for washing DNA or RNA blots are: two washes of twenty minutes each at 20°C to 30°C in 3.0 X SSC. Preferably, moderate stringency wash conditions are utilized after

hybridization in lower stringency hybridization conditions, *i.e.*, 6.0 X SSC, 1% sodium dodecyl sulfate, at 20°C to 30°C for 16 hours.

As used herein, the term "polynucleotide" refers to a deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) molecule that can optionally include one or more non-native nucleotides, having, for example, one or more modifications to the base, sugar, or phosphate portion, or can include a modified phosphodiester linkage. The term polynucleotide includes both single-stranded and double-stranded polynucleotide molecules, which can represent the sense strand, anti-sense strand, or both, and includes linear, circular and branched conformations. Exemplary polynucleotides include genomic DNA, cDNA, mRNA and oligonucleotides, corresponding to either the coding or non-coding portion of the molecule. A polynucleotide of the invention can additionally contain, if desired, a detectable moiety such as a radiolabel, fluorochrome, ferromagnetic substance, luminescent tag or a detectable agent such as biotin.

As used herein, the term "isolated" in regard to a polynucleotide of the invention, is intended to mean a polynucleotide whose structure is not identical to that of any naturally occurring polynucleotide or to that of any fragment of a naturally occurring genomic polynucleotide spanning more than three separate genes. The term therefore includes, for example, (a) a DNA which has the sequence of part of a naturally occurring genomic DNA molecule but is not flanked by both of the coding sequences that flank that part of the molecule in the genome of the organism in which it naturally occurs; (b) a polynucleotide incorporated into a vector or into the genomic DNA of a prokaryote or eukaryote in a manner such that the resulting molecule is not identical to any naturally occurring extrachromosomally replicating DNA or genomic DNA; (c) a separate molecule such as a cDNA, a genomic fragment, a fragment produced by polymerase chain reaction (PCR), or a restriction endonuclease polynucleotide fragment; and (d) a recombinant nucleotide sequence that is part of a hybrid gene, i.e., a gene encoding a fusion protein. Specifically excluded from this definition are polynucleotides present in mixtures of (i) DNA molecules, (ii) transfected cells, and (iii) cell clones, e.g., as these occur in a DNA library such as a cDNA or genomic library.

As used herein, the term "isolated" in regard to a polypeptide of the invention, is intended to mean a molecule that is substantially free from cellular components or other contaminants that are associated with the molecule as it is found in nature. "Substantially pure" or "substantially free" means, in one illustrative aspect of the invention, purified to a purity level of about 85%. In other aspects, these terms denote a purity of at least 90%. In yet other aspects, these terms refer to a purity level of at least 95%. A substantially pure polynucleotide or polypeptide will generally resolve as a band by gel electrophoresis, and generate a nucleotide or amino acid sequence profile consistent with a predominant species.

As used herein, the terms "amino acid" and "amino acids" refer to all naturally occurring L-α-amino acids or their residues. The amino acids are identified by either the single-letter or three-letter designations:

Asp	D	aspartic acid	Ile	I	isoleucine
Thr	T	threonine	Leu	L	leucine
Scr	S	serine	Tyr	Y	tyrosine
Glu	E	glutamic acid	Phe	F	phenylalanine
Pro	P	proline	IIis	II	histidine
Gly	G	glycine	Lys	K	lysine
Ala	A	alanine	Arg	R	arginine
Cys	C	cysteine	Trp	W	tryptophan
Val	V	valine	Gln	Q	glutamine
Met	M	methionine	Asn	N	asparagine

As used herein, the term "ARSDR1" refers to a polypeptide termed androgen regulated short-chain dehydrogenase/reductase 1, which has substantially the same amino acid sequence as shown in SEQ ID NO:2. ARSDR1 is a member of the short-chain dehydrogenase/reductase superfamily and is predominantly expressed in normal and neoplastic prostate epithelium. The ARSDR1 polypeptide is encoded by an approximately 2.5 kb message having the nucleic acid sequence represented in SEQ ID NO:1. The ARSDR1 promoter and regulatory region is approximately 3.1 kb in size and has the sequence shown as nucleotides 1 to 3,113 of SEQ ID NO:8 (genomic nucleotide sequence of ARSDR1). The ARSDR1 promoter contains an androgen response element (ARE) at nucleotides 2,246 to 2,2559 of SEQ ID NO:8 (Roche et al., *Mol. Endocrinol.* **6**: 2229-2235 (1992)) as well as two progesterone responsive elements (PREs) at positions 2,175 to 2,189 and 2,627 to 2,641 in SEQ ID NOS:8 (Lieberman et al., *Mol. Endocrinol.* 7: 515-527(1993)).

As used herein, the term "TMPRSS2" is intended to refer to a polypeptide having substantially the same amino acid sequence as presented in SEQ ID NO:4. The TMPRSS2 polypeptide sequence was also previously described by Paoloni-Giacobino et al., *Genomics 44*:309-329 (1997). Briefly, TMPRSS2 is an androgen-regulated serine protease expressed in normal and neoplastic prostate epithelium. The TMPRSS2 polypeptide is encoded by an approximately 3.8 kb message having the nucleic acid sequence shown in SEQ ID NO:3. The TMPRSS2 promoter and regulatory region is approximately 0.9 kb in size and has the nucleotide sequence presented in SEQ ID NO:9. The TMPRSS2 promoter region contains an androgen response element (ARE) at nucleotides 576 to 590 of SEQ ID NO:9.

As used herein, the term "PART-1" refers to a polypeptide termed prostate androgen-regulated transcript, which has substantially the same amino acid sequence as

shown in SEQ ID NO:6. PART-1 is encoded by an androgen-regulated cDNA whose nucleotide sequence is represented in SEQ ID NO:5. The PART-1 polypeptide is encoded by an approximately 2.1 kb message. The promoter and regulatory region of the polynucleotide encoding PART-1 is contained in an about 2 kb base pair region having the sequence shown in SEQ ID NO:11. The PART-1 promoter region contains a putative binding site for the homeo-domain containing protein Pbx-1a (Van Dijk et al., *Proc. Natl. Acad. Sci* 90:6061-6065 (1993)) at nucleotides 536 to 544 of SEQ ID NO:11.

As used herein, the term "fragment" as used in reference to a substantially pure polynucleotide of the present invention is intended to refer to a portion of the polynucleotide molecule having the ability to selectively hybridize with the parent polynucleotide molecule. The term "selectively hybridize" refers to an ability to bind the parent polynucleotide molecule without substantial cross-reactivity with a molecule that is not the parent polynucleotide molecule. Therefore, the term includes specific hybridization where there is little or no detectable cross-reactivity with other polynucleotide molecules. The term also includes minor cross-reactivity with other molecules provided hybridization to the subject polynucleotide molecule is distinguishable from hybridization to the cross-reactive species. Thus, a fragment of a polynucleotide of the invention can be used, for example, as a PCR primer to selectively amplify a nucleic acid molecule of the invention; as a selective primer for 5' or 3' RACE to determine additional 5' or 3' sequence of a polynucleotide molecule of the invention; as a selective probe to identify or isolate a polynucleotide of the invention on a RNA or DNA blot, or genomic or cDNA library; or as a selective inhibitor of transcription or translation of an inventive polynucleotide in a tissue, cell or cell extract.

The following GenBank Expressed Sequence Tags are specifically excluded as fragments of the invention:

1) ARSDR1 related fragments: (EST) AA 035790, AA 442517, AA 587226, AA 454187, AI 659469, AA 076597, AA 828243, AI 753763, AI 051146. Also excluded as a fragment of the invention is the BAC clone R-1012A1 (GenBank accession number: AL 049779).

2) TMPRSS2 related fragments: (EST) AI 393270, AA 60224, PN_10D11_bd.r1, AI 660243, AI 674580, AA 225818, AA 534046, D25996, AA 876896. Also excluded as a fragment of the invention is the 216 bp nucleic acid described in Paoloni-Giacobino et al., *supra*.

3) PART-1 related fragments: (EST) AA 410580, AA 640889, AI 627693, AI 269149, AA 419011, AA 569503, AI 870129, AA 226501, AA 226220 .

A fragment of a polynucleotide molecule of the invention includes at least about 15 contiguous nucleotides from the reference polynucleotide or a complementary sequence thereto, can include at least about 16, 17, 18, 19, 20 or at least 25 nucleotides,

often includes at least about 30, 40, 50, 100, 300 or 500 nucleotides, and can include up to the full length of the reference polynucleotide molecule minus one nucleotide Fragments of such lengths are able to selectively hybridize with the subject polynucleotide in a variety of detection formats described herein.

As used herein, the term "functional fragment," when used in reference to a polynucleotide comprising the ARSDR1 polynucleotide (SEQ ID NO:8), is intended to refer to any portion of the ARSDR1 polynucleotide having at least one of the biological activities of the subject polynucleotides. Thus, a functional fragment can be a portion of the polynucleotide that enhances or suppresses transcription. For example, a functional fragment of the ARSDR1 polynucleotide (SEQ ID NO:8) may contain an androgen response element (ARE) located at nucleotides 2,246 to 2,259 of SEQ ID NO:8 that exhibits increased expression upon androgen exposure. Alternatively, a functional fragment of the ARSDR1 polynucleotide may contain a progesterone response element (PRE) located at nucleotides 2,175 to 2,189 of SEQ ID NO:8 and nucleotides 2,627 to 2,641 of SEQ ID NO:8, respectively, relative to the transcription start site that exhibits increased expression upon progesterone exposure.

As used herein, the term "functional fragment" when used in reference to the 5' promoter and regulatory region of TMPRSS2 (SEQ ID NO: 9) is intended to refer to a portion of SEQ ID NO:10 having at least one of the activities of its parent polynucleotide molecule. For example, a functional fragment of SEQ ID NO:9 may contain an ARE located at nucleotides 576-590 of SEQ ID NO:9 that exhibits increased expression upon androgen exposure.

As used herein, the term "functional fragment" when used in reference to a polypeptide of the invention, is intended to refer to a peptide fragment that is a portion of a full length polypeptide, provided that the portion has a biological activity that is characteristic of the corresponding full length polypeptide. The term is also intended to include polypeptides that include, for example, modified forms of naturally occurring amino acids such as D-steroisomers, non-naturally occurring amino acids, amino acid analogues and mimetics so long as such polypeptides retain functional activity as defined below.

More specifically, the term "functional fragment" when used in reference to anARSDR1 polypeptide, refers to any peptide sequence which can be identified using the binding and routine methods, such as bioassays described herein. An ARSDR1 polypeptide functional fragment can be, for example, a NAD(H)/NADP(H) binding site referenced herein as amino acids 44 to 50 of SEQ ID NO:2 or a catalytic activity site referenced as amino acids 198 to 202 of SEQ ID NO:2.

As used herein, the term "functional fragment" when used in reference to a PART-1, or TMPRSS2 polypeptide is intended to refer to a portion of the polypeptide

which retains some or all of prostate-specifity and androgen regulated expression of the full length polypeptides shown in SEQ ID NOS:4 and 6.

The term "substantially the nucleotide sequence," as used herein in reference to a polynucleotide of the invention, is intended to mean one of the sequences shown as SEQ ID NOS:1, 3, 5, 7, 8, 9, 10, and 11 or a similar, non-identical sequence that is considered by those skilled in the art to be a functionally equivalent sequence. For example, a polynucleotide sequence that has one or more nucleotide additions, deletions or substitutions with respect to the subject polynucleotide is encompassed by the invention, so long as the polynucleotide sequence encodes the same amino acid sequence or retains its ability to selectively hybridize with the subject polynucleotide. A polynucleotide having substantially the sequence of one of the subject polynucleotides can encode, for example, an isotype variant or species homolog. In addition, a polynucleotide having substantially the nucleotide sequence of the reference polynucleotide has at least 60% identity with respect to the reference nucleotide sequence. A polynucleotide having substantially the same nucleotide sequence of the reference polynucleotide can have at least 70%, at least 90%, or at least 95% identity to the reference nucleotide sequence.

Sequence comparisons between two (or more) polynucleotides or polypeptides are typically performed by comparing sequences of the two sequences over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window", as used herein, refers to a segment of at least about 20 contiguous positions, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned.

Optimal alignment of sequences for comparison may be conducted by local identity or similarity algorithms such as those described in Smith and Waterman (*Adv. Appl. Math.* 2:482 (1981)), by the homology alignment algorithm of Needleman and Wunsch, (*J. Mol. Biol.* 48:443 (1970)), by the search for similarity method of Pearson and Lipman, (*Proc. Natl. Acad. Sci.* 85:2444 (1988)), or by the algorithm of Karlin and Altschul (*Proc. Natl. Acad. Sci.* 87:2264-2268 (1990); *Proc. Natl. Acad. Sci.* 90:5873-5877 (1990)). Computerized implementations of these algorithms are commonly used in the art, such as: GAP, BESTFIT, BLAST, BLASTP2.0.9, TBLASTN, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, Wis.; Atlschul et al., 1997), or NBLAST, and XBLAST (Altschul et al., *J. Mol. Biol.* 215:403-410, (1990)). See also http://www.ncbi.nlm.nih.gov. To obtain gapped alignments for comparison purposes, Gapped BLAST is nutilized as described in Altschul et al. (*Nucleic Acids Res.* 25:3389-3402 (1997).

The term "percent identity" means the percentage of amino acids or nucleotides that occupy the same relative position when two amino acid sequences, or two nucleic acid sequences are aligned side by side using the BLAST programs available at http://www.ncbi.nlm.nih.gov. BLAST nucleotide searches are performed with the NBLAST program, score = 100, wordlength = 12, to obtain nucleotide sequence identity. BLAST protein searches are performed with the XBLAST program, score = 50, wordlength = 3, to obtain amino acid sequence identity. To obtain gapped alignments for comparison purposes, Gapped BLAST is utilized as described in Altschul et al., (*Nucleic Acids Res.* 25:3389-3402 (1997)). When utilitzing BLAST and Gapped BLAST programs, the default parameter of the respective programs (e.g., XBLAST and NBLAST) are used. Neither N- or C- terminal extensions nor insertions shall be construed as reducing sequence identity. See http://www.ncbi.nlm.nih.gov.

The term "percent similarity" is a statistical measure of the degree of relatedness of two compared protein sequences. The percent similarity is calculated by a computer program that assigns a numerical value to each compared pair of amino acids based on chemical similarity (e.g., whether the compared amino acids are acidic, basic, hydrophobic, aromatic, etc.) and/or evolutionary distance as measured by the minimum number of base pair changes that would be required to convert a codon encoding one member of a pair of compared amino acids to a codon encoding the other member of the pair. Calculations are made after a best fit alignment of the two sequences have been made empirically by iterative comparison of all possible alignments. (Henikoff et al., 1992 *Proc. Natl. Acad. Sci. USA* 89:10915-10919).

The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 60% sequence identity, preferably at least 70%, more preferably at least 80% and most preferably at least 90%, compared to a reference sequence using the programs described above (preferably BLAST) using standard parameters. One of skill will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 60%, preferably at least 70%, more preferably at least 80%. Polypeptides which are "substantially similar" share sequences as noted above except that residue positions which are not identical may differ by conservative amino acid changes. Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a

group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine.

As used herein, the term "substantially the amino acid sequence" when used in reference to a TMPRSS2 polypeptide is intended to refer to any amino acid sequence having at least about 56% identity with respect to the reference amino acid sequence shown as SEQ ID NO:4. A polypeptide having substantially the same amino acid sequence as the reference polypeptide can have, for example, 60%, 70%, 80%, 90% or more amino acid sequence identity to the reference amino acid sequence shown as SEQ ID NO:4. Amino acid sequence identity can be determined, for example, in the following manner. The portion of the amino acid sequence of TMPRSS2 (SEQ ID NO:4) extending from amino acid 1 up to and including amino acid 492 is used to search a nucleic acid sequence database, such as the Genbank database, using the program BLASTP version 2.0.9 (Altschul et al., 1997 *Nucleic Acids Res.* 25:3389-3402).

As used herein, the term "substantially the amino acid sequence" when used in reference to an ARSDR1 polypeptide is intended to refer to any amino acid sequence having at least about 26% identity with respect to the reference amino acid sequence shown as SEQ ID NO:2. A polypeptide having substantially the same amino acid sequence as the reference polypeptide can have, for example, 30%, 40%, 50%, 60%, 70%, 80%, 90% or more amino acid sequence identity to the reference amino acid sequence shown as SEQ ID NO:2.

A polypeptide having substantially the amino acid sequence of the reference polypeptide retains comparable functional and biological activity characteristic of the reference polypeptide. It is recognized, however, that polypeptides, or encoding nucleic acids, containing less than the described levels of sequence identity arising as splice variants or that are modified by conservative amino acid substitutions are also encompassed within the scope of the present invention.

As used herein, the term "probe" is intended to refer to a single-stranded polynucleotide, or analogs thereof, that has a sequence of nucleotides that includes at least 10, at least 20, at least 50, at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous bases that are the same as or the complement of any contiguous bases set forth in any of SEQ ID NOS:1, 3, 5, 7, 8, 9, 10 and 11. In addition, the entire sequence corresponding to SEQ ID NOS:1, 3, 5, 7, 8, 9, 10 and 11 can be used as a probe. A probe has the ability to selectively hybridize to its subject polynucleotide

molecule and can be labeled by methods well-known in the art, as described hereinafter, and used, for example, in various diagnostic kits.

The term "antibody" encompasses polyclonal and monoclonal antibody preparations, as well as preparations including hybrid antibodies, altered antibodies, $F(ab')_2$ fragments, F(ab) molecules, Fv fragments, single domain antibodies, chimeric antibodies and functional fragments thereof which exhibit immunological binding properties of the parent antibody molecule.

As used herein, the term "monoclonal antibody" refers to an antibody composition having a homogeneous antibody population. The term is not limited by the manner in which it is made. The term encompasses whole immunoglobulin molecules, as well as Fab molecules, $F(ab')_2$ fragments, Fv fragments, and other molecules that exhibit immunological binding properties of the parent monoclonal antibody molecule. Methods of making polyclonal and monoclonal antibodies are known in the art and described more fully below.

The term "antigen" is defined herein to include any substance that may be specifically bound by an antibody molecule. An "immunogen" is an antigen that is capable of initiating lymphocyte activation resulting in an antigen-specific immune response.

The term "epitope" is used herein to mean a site on an antigen to which specific B-cells and T-cells respond. The term is also used interchangeably with "antigenic determinant" or "antigenic determinant site." A peptide epitope can comprise 3 or more amino acids in a spatial conformation unique to the epitope. Generally, an epitope consists of at least 5 such amino acids and, more usually, consists of at least 8-10 such amino acids. Methods of determining spatial conformation of amino acids are known in the art and include, for example, x-ray crystallography and 2-dimensional nuclear magnetic resonance spectroscopy. Furthermore, the identification of epitopes in a given protein is readily accomplished using techniques well known in the art. See, e.g., Geysen et al. *Proc. Natl. Acad. Sci.* 81:3998 (1984)(general method of rapidly synthesizing peptides to determine the location of immunogenic epitopes in a given antigen); U.S. Pat. No. 4,708,871 (procedures for identifying and chemically synthesizing epitopes of antigens); and Geysen et al. *Molecular Immunology* **23**:709 (1986)(technique for identifying peptides with high affinity for a given antibody). Antibodies that recognize the same epitope can be identified in a simple immunoassay showing the ability of one antibody to block the binding of another antibody to a target antigen.

A "standard" as used herein is a quantitative or qualitative measurement of a compound at a known concentration used for comparing samples with unknown concentrations of the same or related compounds. Preferably, it is based on a statistically appropriate number of samples and is created to use as a basis of comparison when

performing diagnostic assays. Diagnostic assays may in turn be used for monitoring clinical trials, or following patient treatment profiles.

As described herein, the term "prostate neoplastic condition" is intended to refer to a benign or malignant and metastatic prostate lesion of proliferating cells. For example, primary prostate tumors are classified into stages TX, T0, T1, T2, T3, and T4. Metastatic prostate cancer is classified into stages D1, D2, and D3. The term is also intended to include prostate neoplasma.

As used herein, the term "sample" is intended to mean any biological fluid, cell, tissue, organ or portion thereof, that includes or potentially includes nucleic acids and polypeptides of the invention. The term includes samples present in an individual as well as samples obtained or derived from the individual. For example, a sample can be a histologic section of a specimen obtained by biopsy, or cells that are placed in or adapted to tissue culture. A sample further can be a subcellular fraction or extract, or a crude or substantially pure nucleic acid or protein preparation. A sample can be prepared by methods known in the art suitable for the particular format of the detection method.

As used herein, the term "detectable label" refers to a molecule that renders a nucleic acid of the invention detectable by an analytical method. An appropriate detectable label depends on the particular assay format and are well known by those skilled in the art. For example, a detectable label specific for a polynucleotide molecule can be a complementary polynucleotide molecule, such as a hybridization probe, that selectively hybridizes to the polynucleotide molecule. A hybridization probe can be labeled with a measurable moiety, such as a radioisotope, fluorochrome, chemiluminescent marker, biotin, or other moiety known in the art that is measurable by analytical methods. A detectable label also can be a polynucleotide molecule without a measurable moiety. For example, PCR or RT-PCR primers can be used without conjugation to selectively amplify all or a desired portion of the polynucleotide molecule. The amplified polynucleotide can then be detected by methods known in the art.

As used herein, the term "binding agent" when used in reference to ARSDR1, TMPRSS2, and PART-1 polypeptides is intended to mean a compound, a macromolecule, including polypeptide, DNA, RNA and carbohydrate that selectively binds a reference polypeptide or fragment thereof. For example, a binding agent can be a polypeptide that selectively binds with high affinity or avidity to the polypeptides of the present invention, without substantial cross-reactivity with other polypeptides that are unrelated to the reference polypeptide. The affinity of a binding agent that selectively binds to a reference polypeptide will generally be greater than about 10^{-5} M and more usually greater than about 10^{-6} M. High affinity interactions can be preferred, and will generally be greater than about 10^{-8} M to 10^{-9} M. Specific samples of such selective binding agents include a polyclonal or monoclonal antibody specific or selective for a

polypeptide of the present invention or a peptide, polynucleotide, nucleic acid, nucleic acid derivative, steroid or steroid analog, small organic molecule, identified, for example, by affinity screening of a library. For certain applications, a binding agent can be utilized that preferentially recognizes a particular conformational or post-translationally modified state of ARSDR1, TMPRSS2, or PART-1 polypeptides. The binding agent can be labeled with a detectable moiety, if desired, or rendered detectable by specific binding to a detectable secondary binding agent.

As used herein, the term "expression level" when used in reference to ARSDR1, TMPRSS2, and PART-1 is intended to refer to the extent, amount or rate of synthesis of the nucleotide sequences shown as SEQ ID NOS:1, 3, 5, 8, 9, 10 and 11 or the polypeptides shown as SEQ ID NOS: 2, 4 and 6. The extent, amount or rate of synthesis can be determined by measuring the accumulation or synthesis of the reference RNA, reference polypeptide or by measuring the reference polypeptide activity.

As used herein, the term "analog" when used in reference to a short-chain dehydrogenase/reductase substrate is intended to mean any agent which can be oxidized or reduced in the presence of ARSDR1. For example, the short-chain dehydrogenase/reductase substrate analog can be a heterocyclic organic compound having minor modifications of the short-chain dehydrogenase/reductase substrate amino acid sequence. Within the biological arts, the term "about" when used in reference to a particular activity or measurement is intended to refer to the referenced activity or measurement as being within a range of values encompassing the referenced value and within accepted standards of a credible assay within the art, or within accepted statistical variance of a credible assay within the art.

As used herein, the term "analog" when used in reference to a serine protease substrate is intended to mean any agent which is cleaved at about the same rate in the presence of TMPRSS2 as the referenced polypeptide. For example, the serine protease substrate analog can be a peptide having minor modifications of the serine protease substrate amino acid sequence.

As used herein, the term "inhibitor" when used in reference to ARSDR1 is intended to refer to an agent effecting a decrease in the extent, amount or rate of ARSDR1 expression or effecting a decrease in the activity of ARSDR1. For example, one group of inhibitors which decrease the activity of ARSDR1, include short-chain dehydrogenase/reductase inhibitors. Specific examples of short-chain dehydrogenase/reductase inhibitors include, for example, steroids, steroid derivatives and analogs. Other examples of ARSDR1 inhibitors which effect a decrease in ARSDR1 expression include ARSDR1 antisense polynucleotides and transcriptional inhibitors that bind to the ARSDR1 5' promoter and regulatory region.

As used herein, the term "inhibitor" when used in reference to TMPRSS2 is intended to refer to an agent effecting a decrease in the extent, amount or rate of TMPRSS2 expression or effecting a decrease in the activity of TMPRSS2 activity. For example, one group of inhibitors which decrease the activity of TMPRSS2, include serine protease inhibitors. Specific examples of serine protease inhibitors include, for example, antitrypsin and antithrombin. Examples of TMPRSS2 inhibitors which effect a decrease in TMPRSS2 expression include TMPRSS2 antisense polynucleotides and transcriptional inhibitors that bind to the TMPRSS2 5' promoter/regulatory region.

As used herein, the term "inhibitory amount" is intended to refer to the amount of an inhibitor necessary to effect a reduction of at least about 2-fold in the extent, amount or rate of transcription and/or protein synthesis and/or activity.

As used herein, the term "reduced coenzyme" when used in reference to ARSDR1 is intended to refer to a coenzyme that has been reduced during a dehydrogenation reaction mediated by ARSDR1. During the dehydrogenation reaction the substrate is oxidized by the removal of two hydrogen atoms from the substrate. One of the removed hydrogen atoms is directly transferred to the coenzyme, thereby reducing the coenzyme, for example, nicotinamide-adenine dinucleotide (NAD^-) to NADH or nicotinamide-adenine dinucleotide phosphate ($NADP^-$) to NADPH. As used herein, the term "non-reduced coenzyme" is intended to refers to the ARSDR1 coenzyme in its oxidized form, for example, NAD^+ or $NADP^+$.

As used herein, the term "substrate" when used in reference to ARSDR1 is intended to refer to the non-oxidized state of a reactant that is known to become oxidized in an ARSDR1-catalyzed reaction. The term "product," when used in reference to ARSDR1 as used herein is intended to refer to a reactant in an oxidized state that is the product of a dehydrogenation reaction catalyzed by ARSDR1.

Isolation of Prostate-Specific cDNA Clones

The androgen-regulated prostate specific polynucleotide molecules of the present invention were identified by hybridization screening of prostate mRNA against a diverse population of prostate derived probes which were immobilized in a two-dimensional array. A complete description of the methods used for identification, cloning and sequencing of transcripts (SEQ ID NOs:1, 3, 5, and 7) are set forth in the Example sections corresponding to each of the referenced polynucleotides.

In brief, two-dimensional microarrays containing a diverse set of prostate derived cDNAs were screened using RNA from a prostate cell line. A non-redundant set of 1500 prostate-derived cDNA clones was identified from the Prostate Expression Database. The inserts of the cDNAs were amplified with primers BL_ml3F (5'-GTAAAACGACGGCCAGTGAATTG-3') (SEQ ID NO:12) and BL_ml3R (5'-ACACAGGAAACAGCTATGACCATG-3') (SEQ ID NO:13) utilizing PCR and

spotted onto glass microscope slides to form a microarray. To identify genes transcriptionally regulated by androgens, the microarrays of prostate derived cDNAs were screened using total RNA isolated from LNCap cells cultured for 72 hours either in the presence or absence of androgen.

Hybridized microarray slides were scanned with an Array Scanner Generation II (Amersham, Piscataway, NJ). Intensity ratios for each cDNA clone hybridized with probes derived from androgen-stimulated LNCaP and androgen-starved LNCaP were calculated (stimulated intensity/starved intensity). A gene expression level change was treated as significantly different between the two conditions if all four replicate spots for a given cDNA demonstrated a ratio greater than 2 or less than ½ and the signal intensity was greater than 2 standard deviations above the image background.

Microarray hybridization with androgen-stimulated and androgen-starved LNCaP cDNA probes revealed four cDNAs, designated ARSDR1, TMPRSS2, PART-1 and clone 8C3 whose expression was consistently up-regulated using the above criteria. Sequence analysis and BLAST searches against the GenBank databases identified ARSDR1, PART-1 and 8C3 cDNAs as novel genes. Sequence database analysis of the TMPRSS2 cDNA revealed it to be identical to a previously identified serine protease gene that had been mistakenly designated as by expressed in a small intestine-specific fashion (Paoloni-Giacobino et al., *Genomics* 44:309-329 (1997)). The polynucleotide and polypeptide sequences of the present inventive are generally described in the following sections corresponding to each of the prostate-specific, androgen-regulated polynucleotides of the present invention.

ARSDR1

ARSDR1 is a multidomain short-chain dehydrogenase/reductase that is androgen-regulated and predominantly expressed in prostate tissue. ARSDR1 is predominantly transcribed into an about 2.5 kb mRNA transcript. A polynucleotide corresponding to the 5' promoter and regulatory region of the ARSDR1 transcript was identified (Example 5). The ARSDR1 promoter and regulatory region is about 3113 nucleotides in length and is set forth as nucleotides 1 to 3113 of SEQ ID NO:8. As described herein, ARSDR1 was identified as a prostate specific and androgen regulated polynucleotide. Consistent with these functional characteristics, the nucleotide sequence of the ARSDR1 promoter and regulatory region was found to include an androgen response element (ARE). The ARE in ARSDR1 is located at nucleotides 2,246 to 2,259 and is an about 15 nucleotide sequence with substantial similarity to consensus AREs (Roche et al., *Mol. Endocrinol.* 6:2229-2235 (1992)) In addition, the nucleotide sequence of the ARSDR1 promoter and regulatory region was found to include two progesterone response elements (PREs) at nucleotides 2,175 to 2,189 and 2,627 to 2,641 of SEQ ID NO:8, respectively.

The PREs are about 15 nucleotides in length with substantial similarity to consensus PREs (Lieberman et al., *Mol. Endorcrinol.* 7:515-527 (1993)).

The promoter and regulatory region contains binding sites for various transcription and related regulatory factors. The domains carrying these binding sites are functional as binding agents in a variety of methods known in the art to inhibit or identify factors which bind to one or more of these domains in a sequence specific manner. These domains are also useful to construct expression vectors which confer, for example, tissue specificity and androgen regulation. That is the ARSDR1 promoter and regulatory region (as well as the other polynucleotide promoter regions of the present invention) can be use to make prostate-specific, androgen regulated expression vectors that comprising the inventive promoter operably linked to a heterologous nucleotide sequenc. Functional fragments of the ARSDR1 promoter and regulatory region which independently exhibit one or more binding activities or other transcriptional activity of the full length sequence are therefore considered functional fragments of the promoter and regulatory region. Specific examples are portions of the ARSDR1 nucleotide sequence containing the ARE sequence, set forth as nucleotides 2,246 to 2,259 in SEQ ID NO:8, and the PRE sequences set forth as nucleotides 2,175 to 2,189 and 2,627 to 2,641 in SEQ ID NO:8.

ARSDR1 is a member of the short-chain dehydrogenase/reductase (SDR) family of proteins. SDR are a large family of NAD(H) or NADP(H) dependent oxidoreductases. Members of the SDR family of proteins include many enzymes involved in steroid metabolism including, for example, estradiol 17-beta-dehydrogenase, human 15-hydroxyprostaglandin dehydrogenase, and 11-beta-hydroxisteroid dehydrogenase (Jornvall et al., *Biochemistry* 34:6003-6013(1995)). Proteins belonging to the SDR family share amino acid residue identities of only 15-30%, indicating early evolutionary divergence.

The ARSDR1 polypeptide consists of about 318 amino acid residues having the sequence shown in SEQ ID NO:2. Two consensus sequences are conserved in the SDR family, the NAD(H) or NADP(H) binding domain, a N-terminal segment GlyXXXGLYXXGly (SEQ ID NO:14), and the catalytic domain, a sequence TyrXXXLys (SEQ ID NO:15). (Jornvall et al., *supra*, 1995; Ghosh et al., *Structure* 2:629-640(1994)). The ARSDR1 polypeptide contains both of these motifs conserved in the SDR family. ARSDR1 also contains two Asn-glycosylation sites at amino acid positions 174 and 198 (SEQ ID NO:2) that are conserved among SDR family proteins. Another characteristic of ARSDR1 is that it contains two protein kinase C phosphorylation sites at amino acid positions 57 and 106 (SEQ ID NO:2).

TMPRSS2

TMPRSS2 is encoded by a transcript of about 4,650 nucleotides in length. A complete description of the methods used for identification, cloning and sequencing of

the full length transcript is set forth below in Examples 8-11. The complete nucleotide sequence of the TMPRSS2 encoding transcript is shown in SEQ ID NO:3 and the deduced amino acid sequence is shown in SEQ ID NO:4. The full length transcript contains a 5' untranslated region (UTR) of 56 nucleotides and a 3' UTR consisting of 3,115 nucleotides.

A partial nucleotide sequence and deduced amino acid seqeuence has been published by Paoloni-Giacobino et al., *supra*. However, prostate specific expression has not previously been reported. The TMPRSS2 encoding polynucleotide sequence described herein extends the Paoloni-Giacobino et al. sequence by about 2,172 nucleotides at the 3' terminus. The new TMPRSS2 3' UTR sequence is shown as nucleotides 914 to 3,118 in SEQ ID NO:10. In cloning the full length transcript for TMPRSS2 a partial sequence was initially obtained from cDNA clone 10D11 which consisted of 2,681 nucleotides in length. Clone 10D11 begins about 286 nucleotides 5' to the translation stop codon of TMPRSS2 and terminates about 723 nucleotides from the 3' end of the full length transcript (see SEQ ID NO:3). Therefore, clone 10D11 contains a region of 1,449 nucleotides at its 3' terminus that was not described previously by Paoloni-Giacobino et al.

In addition to the TMPRSS2 full length transcript and fragments described above, a polynucleotide corresponding to the 5' promoter and regulatory region was additionally isolated and sequenced. The method used for identifying this genomic sequence is described further below in Example 11. The isolated promoter/regulatory region is about 869 nucleotides in length and is set forth in SEQ ID NO:9. As described herein, TMPRSS2 was identified as a prostate specific and androgen regulated polynucleotide. Consistent with these functional characteristics, the nucleotide sequence of the TMPRSS2 promoter/regulatory region was found to include an ARE. The ARE is located at nucleotides 576 to 590 of SEQ ID NO:9 and is an about 15 nucleotide sequence with substantial similarity to consensus AREs.

The promoter/regulatory region contains binding sites for various transcription and related regulatory factors. The domains carrying these binding sites are functional as binding agents in a variety of methods known in the art to inhibit or identify factors which bind to one or more of these domains in a sequence specific manner. These domains are also useful to construct expression vectors which confer, for example, tissue specificity and androgen regulation. Fragments of the TMPRSS2 promoter/regulatory region which independently exhibit one or more binding activities or other transcriptional activity of the full length sequence are therefore considered functional fragments of the promoter/regulatory region. A specific example is a TMPRSS2 polynucleotide fragment containing the ARE sequence set forth as nucleotides 576 to 590 of SEQ ID NO:9.

TMPRSS2 is a multidomain serine protease that is predominantly expressed in prostate tissue. The polypeptide consists of 492 amino acid residues in length and includes functional domains for serine protease activity, a scavenger receptor cysteine-rich domain, a LDL receptor class domain and a transmembrane domain. The serine protease domain extends from amino acid residue Arg255 to the carboxyl-terminus of the deduced polypeptide (SEQ ID NO:4). There is about 45-55% identity with several members of the serine protease family, including for example, human hepsin (GenBank accession number: X07002), human enterokinase (GenBank accession number: P98073) and human kallikrein (hk2) (GenBank accession number: P03952). The TMPRSS2 protease domain is similar to the S1 family of the SA clan of serine-type peptidases as described by Rawlings, N.D., and Barret, A.J., *Methods Enzymol.* **244**:19-61 (1994). Chymotrypsin and trypsin are examples of members of this family of serine proteases.

The TMPRSS2 active site residues have been identified as His296, Asp345, and Ser441 and cleavage specificity has been deduced to hydrolyze peptide bonds after Lys or Arg residues due to the presence of Asp435 at the base of the S1 subsite which binds to the substrate (SEQ ID NO:4). TMPRSS2 contains nine conserved cysteine residues with the intersubunit disulfide bond between Cys758-Cys912 (SEQ ID NO:4) joining the catalytic protease subunit with the non-protease domains of the polypeptide. The amino-terminal Ile residue of the protease domain is included within the peptide sequence Arg-Ile-Val-Gly-Gly (RIVGG), which is characteristic for the proteolytic activator site of many serine protease zymogens (Rawlings and Barrett, *supra*).

TMPRSS2 contains a hydrophobic sequence at amino acids 84-106 of SEQ ID NO:4 that is characteristic of a transmembrane domain (Hofmann, K., and Stoffel, W., *Biol. Chem. Hoppe-Seyler* **847**: 166 (1993). The transmembrane is not preceded by a peptide leader sequence, indicating that TMPRSS2 is a type II integral membrane proteins in which the amino terminus is located on the cytoplasmic side of the membrane (Parks, G.D., and Lamb, R.A., *J. Biol. Chem.* **268**:19101-19109 (1993)).

In addition to the transmembrane domain, TMPRSS2 contains a third region characteristic of a low-density lipoprotein receptor A domain (LDLRA domain). This domain extends from Cys113 to Cys148 in TMPRSS2 (SEQ ID NO:4). A characteristic LDLRA domain is about 40 amino acids long and contains 6 disulfide-bonded cysteines (Südhoff et al., *Science* **228**:815-822 (1985)). These cysteines have been identified in TMPRSS2 as amino acid residues 113, 120, 126, 133, 139, and 148 (SEQ ID NO:4).

Finally, TMPRSS2 also contains a scavenger receptor cysteine-rich domain (SRCR). SRCR domains characteristically are about 100 amino acids long and rich in cysteine (Resnick et al., *Trends Biochem. Sci.* **19**:5-8 (1994)). The SRCR domain of TMPRSS2 corresponds to amino acid residues Val149 to Leu242 (SEQ ID NO:4). The SRCR domain of TMPRSS2 contains a consensus sequence characteristic of group A

SRCR. Proteins with SRCR domains are known to be expressed either on the cell surface or secreted into plasma or other body fluids.

PART-1

PART-1 is an androgen-regulated polypeptide that is predominantly expressed in prostate tissue. The PART-1 polypeptide consists of 60 amino acid residues and has two protein kinase phosphorylation sites as well as one tyrosine kinase phosphorylation site. The PART-1 polypeptide is encoded by an approximately 2.1 kb messenger RNA (SEQ ID NO:2). The PART-1 promoter region (SEQ ID NO:3) is approximately 2.0 kb in size and contains a binding site for the homeo-domain containing protein Pbx-1a at nucleotides 536 to 544 of SEQ ID NO: 12 as described by Van Dijk et al., *Proc. Natl. Acad. Sci.*, 90:6061-6065 (1993). The nucleotide sequence corresponding to the PART-1 cDNA combined with a portion of its 5' promoter and regulatory region has been described as clone 14D7. The nucleotide sequence of this clone therefore is a composite sequence of the about 2,106 nucleotide PART-1 cDNA and the about 603 nucleotide 5' promoter and regulatory region of the PART-1 transcription unit.

Also provided is an isolated PART-1 polynucleotide having the nucleotide sequence shown as SEQ ID NO:2. In addition to the PART-1 full length transcript and fragments described above, a polynucleotide corresponding to the 5' promoter and regulatory region was isolated. The method used for identifying the nucleotide sequence of the 5' promoter and regulatory region is described further below in the Example 15. The isolated promoter/regulatory region polynucleotide is about 1969 nucleotides in length and is set forth in SEQ ID NO:11. As described herein, PART-1 was identified as a prostate-specific and androgen regulated transcript. The PART-1 Pbx-1a binding site region is shown in SEQ ID NO:11 as nucleotides 536 to 544.

The PART-1 promoter/regulatory region contains binding sites for various transcription and related regulatory factors. The domains carrying these binding sites are functional as binding agents in a variety of methods known in the art to inhibit or identify factors which bind to one or more of these domains in a sequence specific manner. These domains are also useful to construct expression vectors which confer, for example, tissue specificity and androgen regulation. Fragments of the PART-1 promoter/regulatory region which independently exhibit one or more binding activities or other transcriptional activity of the full length sequence are therefore considered functional fragments of the promoter/regulatory region. A specific example is a PART-1 nucleic acid fragment containing the Pbx-1a binding site set forth in SEQ ID NO:11 as nucleotides 536 to 544.

8C3

Identification, characterization, cloning, and chromosomal localization of cDNA 8C3 was performed according to essentially the same methods described for

ARSDR1, TMPRSS2 and PART-1 above and in Examples 1-16. 8C3 is an androgen-regulated transcript of about 4,500 nucleotides in length that is predominantly expressed in prostate tissue. The 8C3 polynucleotide was identified by hybridization screening of prostate mRNA against a diverse population of prostate derived probes which were immobilized in a two-dimensional array. A complete description of the methods used for identification, cloning and sequencing of the 8C3 polynucleotide is set forth below in the Example 1. The 8C3 cDNA has been mapped to 16Q24, a region of the human genome that has been shown to experience a high incidence of chromosomal loss in advanced prostate cancer. Therefore, it is likely that 8C3 is a tumor suppressor. The nucleotide sequence of the 8C3 encoding transcript is shown in SEQ ID NO:7.

The ARSDR1, TMPRSS2, 8C3 and PART-1 polynucleotide and polypeptide sequences of the invention are collectively referred to herein as the polynucleotides and polypeptides of the invention, respectively.

In one aspect of the present invention, substantially pure polynucleotides are provided that are capable of hybridizing under stringent conditions to at least 15 contiguous nucleotides of the nucleotide sequences shown as SEQ ID NOS:1, 3, 5 and 7, or complementary sequences thereof.

In another aspect of the invention, substantially pure polynucleotides having substantially the nucleotide sequences shown as SEQ ID NOS:1, 3, 5, 7, 8, 9, 10 and 11, or functional fragments thereof are also provided. Functional fragments of SEQ ID NOS:8, 9 and 11 may contain, for example, a 5' promoter or a transcription regulatory region, such, as for example, an androgen regulatory element. The promoter and regulatory regions of the present invention contain binding sites for various transcription and related regulatory factors. The domains carrying these binding sites are functional as binding agents in a variety of methods known in the art to inhibit or identify factors which bind to one or more of these domains in a sequence specific manner. These domains are also useful to construct expression vectors which confer, for example, tissue specificity and androgen regulation. Functional fragments of the ARSDR1, TMPRSS2 and PART-1 promoter and regulatory regions which independently exhibit one or more binding activities or other transcriptional activity of the full length sequence are therefore considered functional fragments of the inventive promoter and regulatory regions.

All of the polynucleotides described above, and fragments thereof are useful as hybridization probes in diagnostic procedures. The probes can be as long as the full length transcript or as short as about 10-15 nucleotides, and preferably about 15-18 nucleotides. They can correspond to coding region or untranslated region sequence. The particular application and degree of desired specificity will be one consideration well known to those skilled in the art in selecting a probe. For example, if it is desired to detect an mRNA encoding one of the prostate-specific polypeptides of the present

invention or other related species, the user can choose coding sequence probes and low stringency hybridization conditions. Alternatively, using high stringency conditions with the same probe will select only polynucleotides that actually encode the referenced inventive polypeptide. Untranslated region sequences are useful regions to construct probes since there is little evolutionary pressure to conserve non-coding domains. However, probes as small as 15 nucleotides are statistically unique sequences within the human genome. Therefore, fragments of the inventive sequences that are generally of 15 nucleotides or more in length can be constructed from essentially any region of the transcript or promoter and regulatory region and be capable of uniquely hybridizing to ARSDR1, TMPRSS2, PART-1 and 8C3 polynucleotides.

The probes can be produced recombinantly or chemically synthesized using methods well known in the art. Additionally, ARSDR1, TMPRSS2, PART-1 and 8C3 hybridization probes can be labeled with a variety of detectable labels including, for example, radioisotopes, fluorescent tags, reporter enzymes, biotin and other ligands. Such detectable labels can additionally be coupled with, for example, colorimetric or photometric indicator substrate for spectrophotometric detection. Methods for labeling and detecting such probes are well known in the art and can be found described in, for example, Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Press, Plainsview, New York (1989), and Ausubel et al., Current Protocols in Molecular Biology (Supplement 47), John Wiley & Sons, New York (1999).

Therefore, the invention further provides a substantially pure polynucleotide probe having substantially the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 8, 9, 10 and 11, or fragment thereof. A fragment of the above referenced polynucleotide probes having substantially the sequence of SEQ ID NOS:1, 3, 5, 7, 8, 9, 10 and 11 can, for example, be an oligonucleotide of about 15-18 nucleotides in length.

In another aspect, the present invention is directed to isolated prostate-specific polypeptides (such as polypeptides encoded by the polynucleotide molecules of the present invention) that are androgen regulated. The polypeptides of the present invention can be isolated, for example, by incorporating a polynucleotide molecule of the invention (such as a cDNA molecule) into an expression vector, introducing the expression vector into a host cell and expressing the polynucleotide molecule to yield polypeptide. The polypeptide can then be purified by art-recognized means. When a crude polypeptide extract is initially prepared, it may be desirable to include one or more proteinase inhibitors in the extract. Representative examples of proteinase inhibitors include: serine proteinase inhibitors (such as phenylmethylsulfonyl fluoride (PMSF), benzamide, benzamidine HCl, ε-Amino-*n*-caproic acid and aprotinin (Trasylol)); cysteine proteinase inhibitors, such as sodium *p*-hydroxymercuribenzoate; competitive proteinase inhibitors, such as antipain and leupeptin; covalent proteinase inhibitors, such as iodoacetate and

N-ethylmaleimide; aspartate (acidic) proteinase inhibitors, such as pepstatin and diazoacetylnorleucine methyl ester (DAN); metalloproteinase inhibitors, such as EGTA [ethylene glycol bis(β-aminoethyl ether) *N,N,N',N'*-tetraacetic acid], and the chelator 1, 10-phenanthroline.

In another aspect, the present invention is directed to antibodies that bind specifically to the prostate-specific polypeptides ARSDR1, TMPRSS2, PART-1 or polypeptide fragments thereof. By way of representative example, antigen useful for raising antibodies can be prepared in the following manner. A full-length cDNA molecule of the present invention (or a cDNA molecule of the invention that is not full-length, but which includes a coding region encoding an antigenic polypeptide) can be cloned into a plasmid vector, such as a Bluescript plasmid (available from Stratagene, Inc., La Jolla, California). The recombinant vector is then introduced into an *E. coli* strain (such as *E. coli* XL1-Blue, also available from Stratagene, Inc.) and the polypeptide encoded by the cDNA is expressed in *E. coli* and then purified. For example, *E. coli* XL1-Blue harboring a Bluescript vector including a cDNA molecule of interest can be grown overnight at 37°C in LB medium containing 100 μg ampicillin/ml. A 50 μl aliquot of the overnight culture can be used to inoculate 5 ml of fresh LB medium containing ampicillin, and the culture grown at 37°C with vigorous agitation to A_{600} = 0.5 before induction with 1 mM IPTG. After an additional two hours of growth, the suspension is centrifuged (1000 x *g*, 15 min, 4°C), the media removed, and the pelleted cells resuspended in 1 ml of cold buffer that preferably contains 1 mM EDTA and one or more proteinase inhibitors, such as those described herein in connection with the purification of the isolated polypeptides of the present invention. The cells can be disrupted by sonication with a microprobe. The chilled sonicate is cleared by centrifugation and the expressed, recombinant polypeptide purified from the supernatant by art-recognized protein purification techniques, such as those described herein.

Alternatively, polypeptide fragments of the inventive proteins can be prepared using peptide synthesis methods that are well known in the art. The synthetic polypeptide fragment can then be used to prepare antibodies that are specific to any one of the proteins of the present invention. Direct peptide synthesis using solid-phase techniques (Stewart et al., Solid-Phase Peptide Synthesis, W H Freeman Co, San Francisco Calif. (1969); Merrifield, *J. Am. Chem. Soc.* 85:2149-2154 (1963)) is an alternative to recombinant or chimeric peptide production. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Foster City, Calif.) in accordance with the instructions provided by the manufacturer. Additionally the polypeptide sequences of the present invention or any fragment thereof may be mutated during direct synthesis and, if desired, combined using chemical methods with other amino acid sequences. The polypeptides used to induce specific antibodies may

have an amino acid sequence consisting of at least five amino acids and preferably at least 10 amino acids. Short stretches of amino acid sequence may be attached with those of another polypeptide, and the chimeric polypeptide used for antibody production. Alternatively, the polypeptide may be of sufficient length to contain an entire domain for antibody recognition.

Representative examples of art-recognized techniques for purifying, or partially purifying, polypeptides from biological material are exclusion chromatography, ion-exchange chromatography, hydrophobic interaction chromatography, reversed-phase chromatography and immobilized metal affinity chromatography.

Hydrophobic interaction chromatography and reversed-phase chromatography are two separation methods based on the interactions between the hydrophobic moieties of a sample and an insoluble, immobilized hydrophobic group present on the chromatography matrix. In hydrophobic interaction chromatography the matrix is hydrophilic and is substituted with short-chain phenyl or octyl nonpolar groups. The mobile phase is usually an aqueous salt solution. In reversed phase chromatography the matrix is silica that has been substituted with longer *n*-alkyl chains, usually C_8 (octylsilyl) or C_{18} (octadecylsilyl). The matrix is less polar than the mobile phase. The mobile phase is usually a mixture of water and a less polar organic modifier.

Separations on hydrophobic interaction chromatography matrices are usually done in aqueous salt solutions, which generally are nondenaturing conditions. Samples are loaded onto the matrix in a high-salt buffer and elution is by a descending salt gradient. Separations on reversed-phase media are usually done in mixtures of aqueous and organic solvents, which are often denaturing conditions. In the case of polypeptide and/or peptide purification, hydrophobic interaction chromatography depends on surface hydrophobic groups and is carried out under conditions which maintain the integrity of the polypeptide molecule. Reversed-phase chromatography depends on the native hydrophobicity of the polypeptide and is carried out under conditions which expose nearly all hydrophobic groups to the matrix, *i.e.*, denaturing conditions.

Ion-exchange chromatography is designed specifically for the separation of ionic or ionizable compounds. The stationary phase (column matrix material) carries ionizable functional groups, fixed by chemical bonding to the stationary phase. These fixed charges carry a counterion of opposite sign. This counterion is not fixed and can be displaced. Ion-exchange chromatography is named on the basis of the sign of the displaceable charges. Thus, in anion ion-exchange chromatography the fixed charges are positive and in cation ion-exchange chromatography the fixed charges are negative.

Retention of a molecule on an ion-exchange chromatography column involves an electrostatic interaction between the fixed charges and those of the molecule, binding involves replacement of the nonfixed ions by the molecule. Elution, in turn, involves

displacement of the molecule from the fixed charges by a new counterion with a greater affinity for the fixed charges than the molecule, and which then becomes the new, nonfixed ion.

The ability of counterions (salts) to displace molecules bound to fixed charges is a function of the difference in affinities between the fixed charges and the nonfixed charges of both the molecule and the salt. Affinities in turn are affected by several variables, including the magnitude of the net charge of the molecule and the concentration and type of salt used for displacement.

Solid-phase packings used in ion-exchange chromatography include cellulose, dextrans, agarose, and polystyrene. The exchange groups used include DEAE (diethylaminoethyl), a weak base, that will have a net positive charge when ionized and will therefore bind and exchange anions; and CM (carboxymethyl), a weak acid, with a negative charge when ionized that will bind and exchange cations. Another form of weak anion exchanger contains the PEI (polyethyleneimine) functional group. This material, most usually found on thin layer sheets, is useful for binding polypeptides at pH values above their pI. The polystyrene matrix can be obtained with quaternary ammonium functional groups for strong base anion exchange or with sulfonic acid functional groups for strong acid cation exchange. Intermediate and weak ion-exchange materials are also available. Ion-exchange chromatography need not be performed using a column, and can be performed as batch ion-exchange chromatography with the slurry of the stationary phase in a vessel such as a beaker.

Gel filtration is performed using porous beads as the chromatographic support. A column constructed from such beads will have two measurable liquid volumes, the external volume, consisting of the liquid between the beads, and the internal volume, consisting of the liquid within the pores of the beads. Large molecules will equilibrate only with the external volume while small molecules will equilibrate with both the external and internal volumes. A mixture of molecules (such as proteins) is applied in a discrete volume or zone at the top of a gel filtration column and allowed to percolate through the column. The large molecules are excluded from the internal volume and therefore emerge first from the column while the smaller molecules, which can access the internal volume, emerge later. The volume of a conventional matrix used for protein purification is typically 30 to 100 times the volume of the sample to be fractionated. The absorbance of the column effluent can be continuously monitored at a desired wavelength using a flow monitor.

A technique that is often applied to the purification of polypeptides is High Performance Liquid Chromatography (HPLC). HPLC is an advancement in both the operational theory and fabrication of traditional chromatographic systems. HPLC systems for the separation of biological macromolecules vary from the traditional column

chromatographic systems in three ways: (1) the column packing materials are of much greater mechanical strength, (2) the particle size of the column packing materials has been decreased 5- to 10-fold to enhance adsorption-desorption kinetics and diminish bandspreading, and (3) the columns are operated at 10-60 times higher mobile-phase velocity. Thus, by way of non-limiting example, HPLC can utilize exclusion chromatography, ion-exchange chromatography, hydrophobic interaction chromatography, reversed-phase chromatography and immobilized metal affinity chromatography. Art-recognized techniques for the purification of proteins and peptides are set forth in Methods in Enzymology, Vol. 182, Guide to Protein Purification, Murray P. Deutscher, ed (1990). In particular, Section IV, chapter 14, of the Deutscher publication discloses representative techniques for the preparation of protein extracts from plant material.

Methods for preparing monoclonal and polyclonal antibodies are well known to those of ordinary skill in the art and are set forth, for example, in chapters five and six of *Antibodies A Laboratory Manual*, E. Harlow and D. Lane, Cold Spring Harbor Laboratory (1988). In one representative example, polyclonal antibodies specific for a purified protein can be raised in a New Zealand rabbit implanted with a whiffle ball. One μg of protein is injected at intervals directly into the whiffle ball granuloma. A representative injection regime is injections (each of 1 μg protein) at day 1, day 14 and day 35. Granuloma fluid is withdrawn one week prior to the first injection (preimmune serum), and forty days after the final injection (postimmune serum).

An antibody is specific for one of the inventive proteins if it is produced against an epitope of the polypeptide and binds to at least part of the natural or recombinant polypeptide. Antibody production includes not only the stimulation of an immune response by injection into animals, but also analogous processes such as the production of synthetic antibodies, the screening of recombinant immunoglobulin libraries for specific-binding molecules (Orlandi et al., *Proc. Natl. Acad. Sci.* **86**:3833-3837 (1989), or Huse et al. *Science* **256**:1275-1281 (1989)), or the *in vitro* stimulation of lymphocyte populations.

Current technology (Winter and Milstein, *Nature* **349**:293-299 (1991)) provides for a number of highly specific binding reagents based on the principles of antibody formation. These techniques may be adapted to produce molecules which specifically bind to the inventive proteins or fragments thereof. Antibodies or other appropriate molecules generated against a specific immunogenic peptide fragment or oligopeptide can be used in Western analysis, enzyme-linked immunosorbent assays (ELISA) or similar tests to establish the presence of or to quantitate amounts of any one the inventive proteins in normal, diseased, or therapeutically treated cells or tissues. Variations on any procedure known in the art for the measurement of protein can be used in the practice of

the instant invention. Such procedures include but are not limited to competitive and non-competitive assay systems using techniques such as radioimmunoassays, ELISA (enzyme linked immunoabsorbent assay), sandwich immunoassays, agglutination assays, complement fixation assays, immunoradiometric assays, fluorescent immunoassays, protein A immunoassays, immunoelectrophoresis assays and the like.

The invention also provides a method of diagnosing or predicting the susceptibility of a prostate neoplastic condition in an individual suspected of having a neoplastic condition of the prostate. The method comprises:

(a) obtaining a fluid sample from an individual;

(b) determining the expression level of a polypeptide selected from the group of polypeptides whose amino acid sequences are shown in SEQ ID NOS: 2, 4, and 6, or a polynucleotide selected from the group of polynucleotides whose nucleotide sequences are shown in SEQ ID NOS: 1, 3, 5, 7, 8, 9, 10, and 11; and

(c) comparing said determined expression level of polypeptide or polynucleotide to a corresponding polypeptide or polynucleotide expression level from a normal fluid sample wherein said measured expression level for said polypeptide or polynucleotide of 2-fold or more from said fluid sample from said individual as compared to said normal fluid sample indicates the presence of a prostate neoplastic condition.

In another aspect of the invention, a method of diagnosing or predicting the susceptibility of a prostate neoplastic condition in an individual suspected of having a neoplastic condition of the prostate is provided. The method is performed by:

(a) obtaining a prostate cell sample of a individual;

(b) determining the expression level of a polypeptide selected from the group of polypeptides whose amino acid sequences are shown in SEQ ID NOS: 2, 4, and 6, or a polynucleotide selected from the group of polynucleotides whose nucleotide sequences are shown in SEQ ID NOS:1, 3, 5, 7, 8, 9, 10, and 11, and

(c) comparing the determined expression levels of at least one polypeptide or polynucleotide to a corresponding polypeptide or polynucleotide expression level from normal prostate cells or from an androgen-dependent cell line, wherein the measured expression level for said polypeptide or polynucleotide of 2-fold or more from the individual compared to normal prostate cells or from an androgen-dependent cell line indicates the presence of a prostate neoplastic condition.

A prostate neoplastic condition is a benign or malignant prostate lesion of proliferating cells. Prostate neoplastic conditions include, for example, prostate intraepithelial neoplasia (PIN) and prostate cancer. Prostate cancer is an uncontrolled proliferation of prostate cells which can invade and destroy adjacent tissues as well as metastasize. Primary prostate tumors can be classified into stages TX, T0, T1, T2, T3, and T4 and metastate tumors can be classified into stages D1, D2 and D3. Similarly,

there are classifications known by those skilled in the art for the progressive stages of precancerous lesions or PIN. The methods herein are applicable for the diagnosis or treatment of any or all stages of prostate neoplastic conditions.

The methods of the invention are also applicable to prostate pathologies other than neoplastic conditions. Such other pathologies include, for example, benign prostatic hyperplasia (BPH) and prostatitis. BPH is one of the most common diseases in adult males. Histological evidence of BPH has been found in more than 40% of men in their fifties and almost 90% of men in their eighties. The disease results from the accumulation of non-malignant nodules arising in a small region around the proximal segment of the prostatic urethra which leads to an increase in prostate volume. If left untreated, BPH can result in acute and chronic retention of urine, renal failure secondary to obstructive uropathy, serious urinary tract infection and irreversible bladder decompensation. Prostatitis is an infection of the prostate. Other prostate pathologies known to those skilled in the art exist as well and are similarly applicable for diagnosis or treatment using the methods of the invention. Various neoplastic conditions of the prostate as well as prostate pathologies can be found described in, for example, Campbell's Urology, Seventh Edition, W.B. Saunders Company, Philadelphia (1998). Therefore, the methods of the invention are applicable to both prostate neoplastic conditions and prostate pathologies.

The invention provides a method of diagnosing or predicting prostate neoplastic conditions based on the finding of a positive correlation between ARSDR1, TMPRSS2, PART-1 and 8C3 polypeptide or polynucleotide expression in neoplastic cells of the prostate and the degree or extent of the neoplastic condition or pathology. The diagnostic methods of the invention are applicable to numerous prostate neoplastic conditions and pathologies as described above. One consequence of progression into these neoplastic and pathological conditions is an increased expression of at least one of the ARSDR1, TMPRSS2, PART-1 and 8C3 polypeptides or polynucleotides in prostate tissue as well as secretion into the circulatory system and urine. The increase in at least one of ARSDR1, TMPRSS2, PART-1 and 8C3 expression in individuals suffering from a prostate neoplastic condition can be measured by comparing the amount of ARSDR1, TMPRSS2, PART-1 or 8C3 mRNA and/or polypeptide to that found, for example, in normal prostate tissue samples or in normal blood or serum samples. A two-fold or more increase in polypeptide or polynucleotide expression in a prostate cell sample relative to samples obtained from normal prostate cells or from an androgen-dependent cell line is indicative of a prostate neoplastic condition or pathology. Similarly, an increase in at least one of ARSDR1, TMPRSS2, PART-1 and 8C3 polypeptide or polynucleotide expression leading to two-fold or more secretion of polypeptide into the blood or other

circulatory fluids of the individual compared to normal blood or fluid samples also is indicative of a prostate neoplastic condition or pathology.

As a diagnostic indicator, the polypeptide or polynucleotide of the present invention can be used qualitatively to positively identify a prostate neoplastic condition or pathology as described above. Alternatively, the inventive reagents also can be used quantitiatively to determine the degree or susceptibility of a prostate neoplastic condition or pathology. For example, successive increases in the expression levels of at least one of ARSDR1, TMPRSS2, PART-1 or 8C3, including levels of secreted polypeptide in circulating fluids and urine, can be used as a predictive indicator of the degree or severity of a prostate neoplastic condition or pathology because increased expression, leading to a rise in accumulated levels, for example, also positively correlates with increased severity of a neoplastic condition of the prostate. The higher the level of expression of any one of ARSDR1, TMPRSS2, PART-1 or 8C3, the later the stage of the prostate neoplastic condition or pathology. For example, increases in expression levels of two-fold or more compared to a normal sample is indicative of at least prostate neoplasia. The inventive polypeptide or polynucleotide probes also can be used quantitatively to distinguish between pathologies and neoplastic conditions as well as to distinguish between the different types of neoplastic conditions.

Correlative increases can be determined by comparison of expression of at least one of ARSDR1, TMPRSS2, PART-1 or 8C3 from the individual having, or suspected of having a neoplastic condition of the prostate to expression levels of the corresponding polypeptide or polynucleotide from known samples determined to exhibit a prostate neoplastic condition. Alternatively, correlative increases also can be determined by comparison of expression of at least one of ARSDR1, TMPRSS2, PART-1 or 8C3 from the test individual to expression levels of other known markers of prostate cancer such as prostate specific antigen (PSA), glandular kallikrein 2 (hK2) and prostase/PRSS18. These other known markers can be used, for example, as an internal or external standard for correlation of stage-specific expression with increases in expression of any one of ARSDR1, TMPRSS2, PART-1 or 8C3 and severity of the neoplastic or pathological condition. Conversely, a regression in the severity of a prostate neoplastic condition or pathology is followed by a corresponding decrease in expression levels of at least one of ARSDR1, TMPRSS2, PART-1 or 8C3 and can similarly be assessed using the methods described above.

Given the teachings and guidance provided herein, those skilled in the art will know or can determine the stage or severity of a prostate neoplastic condition or pathology based on a determination of expression levels for of at least one of ARSDR1, TMPRSS2, PART-1 or 8C3 polypeptides and/or polynucleotides and using known procedures and marker comparisons other than those described above. For a review of

recognized values for such other marker in normal versus pathological tissues, see for example, Campbell's Urology, Seventh Edition, W.B. Saunders Company, Philadelphia (1998).

Therefore, the invention provides a method for both diagnosing and prognosing a prostate neoplastic condition including prostate cancer and prostate interepithelial neoplasia as well as other prostate pathologies such as BPH and prostatitis.

The use of at least one of ARSDR1, TMPRSS2, PART-1 or 8C3 expression levels in prostate cells, the circulatory system and urine as a diagnostic indicator of a prostate pathology allows for early diagnosis as a predictive indicator when no physiological or pathological symptoms are apparent. The methods are applicable to any males, generally those over age 50, African-American males and males with familial history of prostate neoplastic conditions or pathologies. The diagnostic methods of the invention also are applicable to individuals predicted to be at risk for prostate neoplastic conditions or pathologies by reliable prognostic indicators prior to onset of overt clinical symptoms. All that is necessary is to determine the expression level of at least one of ARSDR1, TMPRSS2, PART-1 or 8C3 in prostate tissue or circulatory or bodily fluid to determine whether there is an increase in these polypeptide or polynucleotide levels in the individual suspected of having a prostate pathology compared to normal individuals. Those skilled in the art will know, or be able to determine, by using routine examinations and practices in the field of medicine, those individuals who are applicable candidates for diagnosis by the methods of the invention.

For example, individuals suspected of having a prostate neoplastic condition or pathology can be identified by exhibiting presenting signs of prostate cancer which include, for example, a palpable nodule (which generally occurs in greater than 50% of the cases), dysuria, cystitis and prostatitis, frequency, urinary retention, or decreased urine stream. Signs of advanced disease include pain, uremia, weight loss and systemic bleeding. Prognostic methods of this invention are applicable to individuals after diagnosis of a prostate neoplastic condition, for example, to monitor improvements or identify residual neoplastic prostate cells using, for example, imaging methods known in the art and which targets at least one of ARSDR1, TMPRSS2, PART-1 or 8C3 polypeptides or polynucleotides.

Therefore, the invention provides a method of predicting the onset of a prostate neoplastic condition or pathology. The method consists of determining increased expression levels of at least one of ARSDR1, TMPRSS2, PART-1 and 8C3 in a prostate cell sample or in fluids from an individual having or suspected of having a prostate neoplastic condition or pathology compared to a sample isolated from a normal individual, where increased expression in the sample indicates the onset of the prostate neoplastic condition or pathology.

The diagnostic methods of the invention are applicable for use with a variety of different types of samples isolated or obtained from an individual having, or suspected of having a prostate neoplastic condition or prostate pathology. For example, samples applicable for use in one or more diagnostic formats of the invention, include tissue and cell samples. A tissue or cell sample can be obtained, for example, by biopsy or surgery. As described below, and depending on the format of the method, the tissue can be used whole or subjected to various methods known in the art to disassociate the sample into smaller pieces, cell aggregates or individual cells. Additionally, when combined with amplification methods such as polymerase chain reaction (PCR), a single prostate cell sample is sufficient for use in diagnostic assays of the invention which employ hybridization detection methods. Similarly, when measuring levels of any one of ARSDR1, TMPRSS2, and PART-1 polypeptide or activity levels, amplification of the signal with enzymatic coupling or photometric enhancement can be employed using only a few or a small number of cells.

Whole tissue obtained from a prostate biopsy or surgery is one example of a prostate cell sample. Whole tissue prostate cell samples can be assayed employing any of the formats described below. For example, the prostate tissue sample can be mounted and hybridized *in situ* with a polynucleotide probe of the present invention. Similar histological formats employing protein detection methods and *in situ* activity assays also can be used to detect polypeptides of the invention in whole tissue prostate cell samples. Polypeptide detection methods include, for example, staining with antibodies specific for at least one of the inventive polypeptides and activity assays which result in the deposition of an ARSDR1, TMPRSS2, or PART-1 end product at the site of enzyme activity in the sample. Such histological methods as well as others are well known to those skilled in the art and are applicable for use in the diagnostic methods of the invention using whole tissue as the source of a prostate cell sample. Methods for preparing and mounting the samples are similarly well known in the art.

Individual prostate cells and cell aggregates from an individual having, or suspected of having a prostate neoplastic condition or pathology is another example of a prostate cell sample which can be analyzed for increased expression of ARSDR1, TMPRSS2, PART-1 or 8C3, polypeptide or polynucleotide or activity. The cells can be grown in culture and analyzed *in situ* using procedures such as those described above. The expression level can be determined by, for example, binding agents specific for ARSDR1, TMPRSS2, or PART-1 polypeptides, or by hybridization to a probe specific to at least one of ARSDR1, TMPRSS2, PART-1 and 8C3 polynucleotides. Other methods for measuring the expression level of the inventive polypeptides or polynucleotides in whole cell samples are known in the art and are similarly applicable in any of the diagnostic formats described below.

The tissue or whole cell prostate cell sample obtained from an individual also can be analyzed for increased expression of at least one of ARSDR1, TMPRSS2, PART-1 or 8C3 by lysing the cell and measuring the expression levels of an inventive polypeptide or polynucleotide in the lysate, a fractionated portion thereof or a purified component thereof using any of diagnostic formats described below. For example, if a hybridization format is used, RNA from one or more of the inventive polynucleotides can be amplified directly from the lysate using PCR, or other amplification procedures well known in the art such as RT-PCR, 5' or 3' RACE to directly measure the expression levels of at least one of ARSDR1, TMPRSS2, PART-1 or 8C3. RNA also can be isolated and probed directly such as by solution hybridization or indirectly by hybridization to immobilized RNA. Similarly, when determining the expression level of the polypeptides of the invention using polypeptide detection or enzyme activity formats, lysates can be assayed directly, or they can be further fractionated to enrich for the inventive polypeptides and their corresponding activities. Numerous other methods applicable for use with various cell fractions are well known to those skilled in the art and can accordingly be used in the methods of the invention.

The prostate tissue or cell sample can be obtained directly from the individual or, alternatively, it can be obtained form other sources for testing. Similarly, the cell sample can be tested when it is freshly isolated or it can be tested following short or prolonged periods of cryopreservation without substantial loss in accuracy or sensitivity. If the sample is to be tested following an indeterminate period of time, it can be obtained and then cryopreserved, or stored at 4°C for short periods of time, for example. An advantage of the diagnostic methods of the invention is that they do not require histological analysis of the sample. As such, the sample can be initially disaggregated, lysed, fractionated or purified and the active component stored for later diagnosis.

The diagnostic methods of the invention are applicable for use with a variety of different types of samples other than prostate cell samples. For example, intracellular polynucleotides and polypeptides of the invention may leak into the extracellular space when a neoplastic prostate condition causes a disruption of the normal prostate architecture. Therefore, the diagnostic methods of the invention are applicable with fluid samples collected from an individual having, or suspected of having a neoplastic condition of the prostate or a prostate pathology.

Fluid samples which can be measured for ARSDR1, TMPRSS2, PART-1 or 8C3 expression levels include, for example, blood, serum, lymph, urine and semen. Other bodily fluids are known to those skilled in the art and are similarly applicable for use as a sample in the diagnostic methods of the invention. One advantage of analyzing fluid samples is that they are readily obtainable, in sufficient quantity, without invasive procedures as required by biopsy and surgery. Analysis of fluid samples such as blood,

serum and urine will generally be in the diagnostic formats described above and below which measure ARSDR1, TMPRSS2, or PART-1 polypeptide levels or activity. As the inventive polypeptides are circulating in bodily fluids, the methods will be similar to those which measure expression levels from cell lysates, fractionated portions thereof or purified components.

Prostate neoplastic conditions and prostate pathologies can be diagnosed, predicted or prognosed by measuring the expression levels of the polynucleotides and polypeptides of the present invention in a prostate cell sample, circulating fluid or other bodily fluid obtained from the individual. As described above, expression levels can be measured by a variety methods known in the art. For example, the expression level of a nucleic acid of the invention can be determined by measuring the amount of an RNA or polypeptide of the invention in a sample from the individual. Alternatively, the expression level of the inventive polypeptides can be determined by measuring the amount of enzyme activity in the sample, the amount of activity being indicative of the expression level of the inventive polynucleotide.

Given the teachings and guidance provided herein, the choice of measuring RNA, polypeptide or activity will be that of the user. Considerations such as the sample type, availability and amount will also influence selection of a particular diagnostic format. For example, if the sample is a prostate cell sample and there is only a small amount available, then diagnostic formats which measure the amount of RNA by, for example, PCR amplification, can be an appropriate choice for determining the expression level of a polynucleotide of the invention. Alternatively, if the sample is a blood sample and the user is analyzing numerous different samples simultaneous, such as in a clinical setting, then a multi sample format, such as an Enzyme Linked Immunoabsorbant Assay (ELISA), which measures the amount of polypeptide can be an appropriate choice for determining the expression level of a polypeptide of the invention. Additionally, polynucleotides of the invention released into bodily fluids from the neoplastic or pathological prostate cells can also be analyzed by, for example, PCR or RT-PCR. Those skilled in the art will know, or can determine which format is amenable for a particular application and which methods or modifications known within the art are compatible with a particular type of format.

Hybridization methods are applicable for measuring the amount of inventive RNA as an indicator of expression levels. There are numerous methods well known in the art for detecting polynucleotides by specific or selective hybridization with a complementary probe. Such methods include both solution hybridization procedures and solid-phase hybridization procedures where the probe or sample is immobilized to a solid support. Descriptions for such methods can be found in, for example, Sambrook et al., *supra*, and in Ausubel et al., *supra*. Specific examples of such methods include PCR and

other amplification methods such as RT-PCR, 5' or 3' RACE, RNase protection, RNA blot, dot blot or other membrane-based technologies, dip stick, pin, ELISA or two-dimensional arrays immobilized onto chips as a solid support. These methods can be performed using either qualitative or quantitative measurements, all of which are well known to those skilled in the art.

PCR or RT-PCR can be used with isolated RNA or crude cell lysate preparations. As described previously, PCR is advantageous when there is little starting material. A further description of PCR methods can be found in, for example, Dieffenbach, C.W., and Dveksler, G.S., PCR Primer: A Laboratory Manual, Cold Spring Harbor Press, Plainsview, New York (1995). Multi sample formats such as an ELISA or two-dimensional array offer the advantage of analyzing numerous, different samples in a single assay. A particular example of a two-dimensional array used in a hybridization format is described further below in the Examples. In contrast, solid-phase dip stick-based methods offer the advantage of being able to rapidly analyze a patient's fluid sample and obtain an immediate result.

Polynucleotide probes useful for measuring the expression level of the polynucleotides of the invention by hybridization include, for example, all of the polynucleotides probes described previously. More specifically, ARSDR1 probes include, for example, polynucleotides corresponding to the entire transcribed region of SEQ ID NO:1 and fragments thereof. Similarly, TMPRSS2, PART-1, and 8C3 probes include, for example, polynucleotides corresponding to the entire polynucleotide sequences designated as SEQ ID NOS:1, 3, 5, 7 and fragments thereof, respectively.

Briefly, for detection by hybridization, the polynucleotides probes of the invention having detectable labels are added to a prostate cell sample or a fluid sample obtained from the individual having, or suspected of having a prostate neoplastic condition or pathology under conditions which allow annealing of the probe to RNA. Such conditions are well known in the art for both solution and solid phase hybridization procedures. Moreover, optimization of hybridization conditions can be performed, if desired, by hybridization of an aliquot of the sample at different temperatures, durations and in different buffer conditions. Such procedures are routine and well known to those skilled. Following annealing, the sample is washed and the signal is measured and compared with a suitable control or standard value. The magnitude of the hybridization signal is directly proportional to the expression levels of the polynucleotide of the invention for which the probe was specific.

A suitable control for comparison can be, for example, the expression level of a polynucleotide of the invention from a prostate cell or a fluid sample obtained from a normal individual. Another suitable control for comparison is a prostate cell line that is androgen-dependent. ARSDR1, TMPRSS2, PART-1, and 8C3 expression levels in cell

lines generally should be determined under androgen depleted growth conditions, as their response to androgen stimulation will be indicative of their respective expression levels in neoplastic cells. The control sample for comparison can be measured simultaneously with one or more test samples or, alternatively, expression levels can be established for a particular type of sample and standardized to internal or external parameters such as polypeptide or polynucleotide content, cell number or mass of tissue. Such standardized control samples can then be directly compared with results obtained from the test sample. An increase of two-fold or more of expression levels of a polynucleotide of the invention indicates the presence of a prostate neoplastic condition or pathology in the tested individual.

The diagnostic procedures described above and below using ANSDR1, TMPRSS2, PART-1, and 8C3 polynucleotide and polypeptide probes can additionally be used in conjunction with other prostate markers, such as prostate specific antigen (PSA), human glandular kallikrein 2 (hk2) and prostase/PRSS18 for simultaneous or independent corroboration of a sample. Moreover, while the diagnositic procedures described above and below describe using ANSDR1, TMPRSS2, PART-1, and 8C3 individually, these markers can also be used in combination. Those skilled in the art will know which markers are applicable for use in conjunction with a polynucleotide or polypeptide of the invention to delineate more specific diagnostic information such as that described above.

Therefore, the invention provides a method of diagnosing or predicting the susceptibility of a prostate neoplastic condition in an individual suspected of having a neoplastic condition of the prostate where the expression level of a polynucleotide of the invention is determined by measuring the amount of its respective RNA. The amount of ANSDR1, TMPRSS2, PART-1, and 8C3 RNA can be determined by hybridization with a polynucleotide probe having substantially the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, or functional fragment thereof, respectively, and wherein the fragment consists of an oligonucleotide of about 15-18 nucleotides in length.

The invention additionally provides a method of diagnosing or predicting the susceptibility of a prostate neoplastic condition in an individual suspected of having a neoplastic condition of the prostate where the expression level of an inventive polypeptide is determined by measuring the amount of polypeptide. The method comprises contacting a cell, a cell lysate, or fractionated sample thereof, from an individual suspected of having a neoplastic condition with a binding agent selective for one of the inventive polypeptides, and determining the amount of selective binding of the agent. The fractionated sample can be a cell lysate or lipid membranes and the binding agent can be an antibody or a non-hydrolizable substrate analog depending upon which inventive polypeptide is being assayed. For example, when the assay is directed to

PART-1 the fraction can be lipid membranes and the selective binding agent can be an antibody. Alternatively, when the assayed polypeptide is ARSDR1, the fractionated sample can be a cell lysate and the binding agent can be an antibody or non-hydrolizable short-chain dehydrogenase/reductase substrate analog.

Essentially all modes of affinity binding assays are applicable for use in determining the amount of a polypeptide of the invention in a sample. Such methods are rapid, efficient and sensitive. Moreover, affinity binding methods are simple and can be adjusted to be performed in a variety of clinical settings and under conditions to suit a variety of particular needs. Affinity binding assays which are known and can be used in the methods of the invention include both soluble and solid phase formats. A specific example of a soluble phase affinity binding assay is immunoprecipitation using an antibody selective for a polypeptide of the invention or other binding agent, such as, for example a steroid or steroid derivative for ARSDR1. Solid phase formats are advantageous for the methods of the invention since they are rapid and can be performed more easily on multiple different samples simultaneously without losing sensitivity or accuracy. Moreover, solid phase affinity binding assays are further amenable to high throughput screening and automation.

Specific examples of solid phase affinity binding assays include immunoaffinity binding assays such as an ELISA and radioimmune assay (RIA). Other solid phase affinity binding assays are known to those skilled in the art and are applicable to the methods of the invention. Although affinity binding assays are generally formatted for use with an antibody that is selective for the analyte or ligand of interest, essentially any binding agent can be alternatively substituted for selectively binding the antibody. Such binding agents include, for example, steroids, steroid derivatives, macromolecules such as polypeptides, peptides, nucleic acids, lipids and sugars as well as small molecule compounds. Other binding agents selective for ARSDR1 and TMPRSS2 include, for example, non-hydrolizable short-chain dehydrogenase/reductase substrate analogs and non-hydrolizable serine protease substrate analogs, respectively. Methods are known in the art for identifying such molecules which bind selectively to a particular analyte or ligand and include, for example, combinatorial libraries. Thus, for a molecule other than an antibody to be used in an affinity binding assay, all that is necessary is for the binding agent to exhibit selective binding activity for the inventive polypeptide.

Various modes of affinity binding formats are similarly known which can be used in the diagnostic methods of the invention. For the purpose of illustration, particular embodiments of such affinity binding assays will be described further in reference to immunoaffinity binding assays. The various modes of affinity binding assays, such as immunoaffinity binding assays, include for example, solid phase ELISA and RIA as well as modifications thereof. Such modifications thereof include, for example, capture

assays and sandwich assays as well as the use of either mode in combination with a competition assay format. The choice of which mode or format of immunoaffinity binding assay to use will depend on the intent of the user. Such methods can be found described in common laboratory manuals such as Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York (1999).

As with the hybridization methods described previously, the diagnostic formats employing affinity binding can be used in conjunction with a variety of detection labels and systems known in the art to quantitate amounts of a polypeptide of the invention in the analyzed sample. Detection systems include the detection of bound polypeptide of the invention by both direct and indirect means. Direct detection methods include labeling of an antibody or binding agent that binds specifically to a polypeptide of the invention. Indirect detection systems include, for example, the use of labeled secondary antibodies and binding agents.

Secondary antibodies, labels and detection systems are well known in the art and can be obtained commercially or by techniques well known in the art. The detectable labels and systems employed with a binding agent that is specific to a polypeptide of the invention should not impair binding of the agent to its cognate inventive polypeptide. Moreover, multiple antibody and label systems can be employed for detecting bound antigen/antibody complexes of the invention to enhance the sensitivity of the binding assay if desired.

As with the hybridization formats described previously, detectable labels can be essentially any label that can be quantitated or measured by analytical methods. Such labels include, for example, enzymes, radioisotopes, fluorochromes as well as chemi- and bioluminescent compounds. Specific examples of enzyme labels include horseradish peroxidase (HRP), alkaline phosphatase (AP), β-galactosidase, urease and luciferase.

A horseradish-peroxidase detection system can be used, for example, with the chromogenic substrate tetramethylbenzidine (TMB), which yields a soluble product in the presence of hydrogen peroxide that is detectable by measuring absorbance at 450 nm. An alkaline phosphatase detection system can be used with the chromogenic substrate *p*-nitrophenyl phosphate, for example, which yields a soluble product readily detectable by measuring absorbance at 405 nm. Similarly, a β-galactosidase detection system can be used with the chromogenic substrate *o*-nitrophenyl-β-D-galactopyranoside (ONPG), which yields a soluble product detectable by measuring absorbance at 410 nm, or a urease detection system can be used with a substrate such as urea-bromocresol purple (Sigma Immunochemicals, St. Louis, MO). Luciferin is the substrate compound for luciferase which emits light following ATP-dependent oxidation.

Fluorochrome detection labels are rendered detectable through the emission of light of ultraviolet or visible wavelength after excitation by light or another energy

source. DAPI, fluorescein, Hoechst 33258, R-phycocyanin, B-phycoerythrin, R-phycoerythrin, rhodamine, Texas red and lissamine are specific examples of fluorochrome detection labels that can be utilized in the affinity binding formats of the invention. Particularly useful fluorochromes include fluorescein and rhodamine.

Chemiluminescent as well as bioluminescent detection labels are convenient for sensitive, non-radioactive detection of the inventive polynucleotides and polypeptides and can be obtained commercially from various sources such as Amersham Lifesciences, Inc. (Arlington Heights, IL).

Radioisotopes can alternatively, be used as detectable labels for use in the binding assays of the invention. Iodine-125 is a specific example of a radioisotope useful for a detectable label.

Signals from detectable labels can be analyzed, for example, using a spectrophotometer to detect color from a chromogenic substrate; a fluorometer to detect fluorescence in the presence of light of a certain wavelength; or a radiation counter to detect radiation, such as a gamma counter for detection of iodine-125. For detection of an enzyme-linked secondary antibody, for example, a quantitative analysis of the amount of bound agent can be made using a spectrophotometer such as an EMAX Microplate Reader (Molecular Devices, Menlo Park, CA) in accordance with the manufacturer's instructions. If desired, the assays of the invention can be automated or performed robotically, and the signal from multiple samples can be detected simultaneously.

The diagnostic formats of the present invention can be forward, reverse or simultaneous as described in U.S. Patent No. 4,376,110 and No. 4,778,751. Separation steps for the various assay formats described herein, including the removal of unbound secondary antibody, can be performed by methods known in the art (Harlow and Lane, *supra*). For example, washing with a suitable buffer can be followed by filtration, aspiration, vacuum or magnetic separation as well as by centrifugation.

A binding agent selective for a polypeptide of the invention also can be utilized in imaging methods that are targeted at prostate cells expressing the nucleic acids of the invention. These imaging techniques will have utility in identification of residual neoplastic cells at the primary site following standard treatments including, for example, radical prostatectomy, radiation or hormone therapy. In addition, imaging techniques that detect neoplastic prostate cells have utility in detecting secondary sites of metastasis. A binding agent specific for one of the polypeptides of the invention can be radiolabeled with, for example, ^{111}indium and infused intravenously as described by Kahn et al. (*Journal of Urology* 152:1952-1955 (1994)). The binding agent selective for a polypeptide of the invention can be, for example, a monoclonal antibody selective for any one of the inventive polypeptides. Imaging can be accomplished by, for example, radioimmunoscintigraphy as described by Kahn et al., *supra*.

The invention additionally provides a method of diagnosing or predicting the susceptibility of a prostate neoplastic condition in an individual suspected of having a neoplastic condition of the prostate where the inventive polypeptide expression level is determined by measuring the amount of ARSDR1 or TMPRSS2 enzyme activity. In the case of ARSDR1, the method comprises contacting a cell, a cell lysate, or fractionated sample thereof, from the individual with a with a short-chain dehydrogenase/reductase substrate selective for ARSDR1, and determining the amount of product formed by ARSDR1. When ARSDR1 activity is used in the method the fractionated sample can be cell lysate. Alternatively, when TMPRSS2 is being assayed the inventive method comprises contacting a cell, a cell lysate, or fractionated sample thereof, from the individual with a serine protease substrate selective for TMPRSS2, and determining the amount of cleavage product produced by TMPRSS2. When TMPRSS2 activity is used in the method the fractionated sample can be lipid membranes.

Another diagnostic format which can be used for determining the expression levels of ARSDR1 and TMPRSS2 is by measuring the activity of ARSDR1 short-chain dehydrogenase/reductase activity and serine protease activity, respectively, in a sample. As with the hybridization and affinity binding formats, activity assays can similarly be performed using essentially identical methods and modes of analysis. Therefore, solution and solid phase modes, including multi sample ELISA, RIA and two-dimensional array procedures are applicable for use in measuring the short-chain dehydrogenase/reductase activity of ARSDR1 and the serine protease activity of TMPRSS2. In the case of ARSDR1, activity can be measured by, for example, incubating a short-chain dehydrogenase/reductase substrate with the sample and determining the amount of product formation from the short-chain dehydrogenase/reductase substrate. When TMPRSS2 activity is being measured, a serine protease substrate is incubated with the sample and the amount of protein cleavage is determined. In either case, the enzyme products can be measured using, for example, any of the detectable labels and detection systems described previously.

When ARSDR1 activity is monitored, the amount of product formed or rate of product formation can be measured either indirectly by measuring the appearance of reduced coenzyme or disappearance of non-reduced coenzyme or, can be measured directly by measuring the appearance of product or disappearance of substrate. The amount of product formation can be measured indirectly by measuring the appearance of reduced coenzyme, for example, NADH or NADPH, indicating that the substrate has been oxidized in the ARSDR1-catalyzed reaction. Conversely, the amount of product formed or rate of product formation can be measured indirectly measuring the disappearance of non-reduced coenzyme, for example, NAD^+ and $NADP^+$, indicating that the coenzyme has been reduced in the ARSDR1 catalyzed reaction. In addition, the

appearance of product and disappearance of substrate can also be used to measure the activity of ARSDR1. The magnitude of product formed will directly correlate with the ARSDR1 activity in the sample and therefore, with the expression levels of ARSDR1 in the sample.

Methods applicable for determining the activity of ARSDR1 in a sample include, for example, determining the presence of short-chain dehydrogenase/reductase substrates such as steroids or steroid derivatives containing hydroxyl groups and short-chain dehydrogenase/reductase coenzymes such as pyridine nucleotides NAD^+ and $NADP^+$ or derivatives thereof. Derivatives can further exhibit the capability of releasing a dye or fluorochrome, for example, upon chemical modification by ARSDR1 such as the oxidation of the substrate or reduction of the coenzyme. The difference in light absorbance between the oxidized and reduced forms of coenzyme is routinely distinguished by spectral measurements well known in the art. For example, NADH and NADPH are characterized by maximal absortion at about 340nm, while the non-reduced forms, NAD^+ and $NADP^+$, absorb maximally at about 260mm. Methods useful for the detection of changes in polarity are useful for measuring the disappearance of substrate and appearance of product and can include, for example, thin layer chromatography (TLC), nuclear magnetic resonance spectroscopy (NMR) and infrared spectroscopy. Short-chain dehydrogenase/reductase substrates, coenzymes and their respective derivatives are well known in the art and are similarly applicable in the methods of the invention for determining ARSDR1 activity in a sample.

Substrates applicable for determining the activity of TMPRSS2 in a sample include, for example, serine protease substrates such as Lys and Arg containing polypeptides and peptides. Specific examples of TMPRSS2 substrates include PSA, hk2, semenogelin, hemoglobin, glucagon, and casein, all of which can be obtained from commercial sources. Peptides of these polypeptides can additionally be used as TMPRSS2 substrates so long as they contain a Lys or Arg residue. In addition, serine protease substrate analogs also can be used for determining the amount of TMPRSS2 activity in a sample. Such analogs can further exhibit the capability of releasing a dye or fluorochrome, for example, upon cleavage by TMPRSS2. A serine protease analog capable of releasing dye is azo dye-impregnated collagen, which is also available commercially. Other serine protease substrates and analogs are well known to those skilled in the art and are similarly applicable in the methods of the invention for determining TMPRSS2 activity in a sample.

The invention further provides a method of identifying a compound that inhibits the activity of an inventive polypeptide. The method consists of contacting a sample containing the inventive polypeptide and an appropriate substrate, with a test compound under conditions that allow product formation from the substrate, and measuring the

amount of the product formation from the substrate. A decrease in the amount of product formation from the inventive polypeptide substrate in the presence of the test compound compared to the absence of the test compound indicates that the compound has inhibitory activity towards the inventive polypeptide activity. Similarly, compounds that increase the activity of an inventive polypeptide also can be identified. A test compound added to a sample containing an inventive polypeptide and an appropriate substrate which increases the amount of product or rate of product formation chemical modification of the substrate compared to the absence of the test compound indicates that the compound increases the activity of the inventive polypeptide. Therefore, the invention provides a method of identifying compounds that modulate the activity of the polypeptides of the present invention. The polypeptide containing sample used for such a method can be serum, prostate tissue, a prostate cell population or a recombinant cell population expressing the inventive polypeptide.

The methods for determining the activity of an inventive polypeptide in a sample described above can also be adapted for screening test compounds to determine their ability to inhibit or increase product formation catalyzed by an inventive polypeptide from its substrates. In such cases, a test compound is added to a reaction system and the effect of the test compound on production of product is observed. Those compounds which inhibit the product formation or rate of product formation are considered as potential antagonists of the inventive polypeptides and further as potential therapeutic agents for treatment of neoplastic conditions of the prostate. Similarly, those compounds which increase the product or rate of product formation are considered as potential agonists of the inventive polypeptides and further as potential therapeutic agents for the treatment of neoplastic conditions of the prostate.

A reaction system for identifying a compound that inhibits or enhances the activity of the inventive polypeptides can be performed using essentially any source of inventive polypeptide activity. Such sources include, for example, a prostate cell sample, lysate or fractionated portion thereof; a bodily fluid such as blood, serum or urine from an individual with a prostate neoplastic condition; a recombinant cell or soluble recombinant source, and an *in vitro* translated source. The source of inventive polypeptide is combined with an appropriate substrate as described above and incubated in the presence or absence of a test inhibitory compound. The reaction rate or extent of the usage of the substrate in the presence of the test compound is compared with that in the absence of the test compound. Those test compounds which provide inhibition of the reaction activity of at least about 50% are considered to be inhibitors of the inventive polypeptides. Similarly, those compounds which increase the reaction activity of two-fold for more are considered to be enhancers of the activity of the inventive polypeptides. Such inhibitors of the inventive polypeptides can then be subjected to further *in vitro* or

in vivo testing to confirm that they inhibit the production of substrates of the inventive polypeptides in cellular and animal models.

Suitable test compounds for the inhibition or enhancement assays can be any substance, molecule, compound, mixture of molecules or compounds, or any other composition which is suspected of being capable of inhibiting inventive polypeptide activity *in vivo* or *in vitro*. The test compounds can be heterocyclic organic compounds such as steroids or steroid derivatives, macromolecules, such as biological polymers, including proteins, polysacchrides and nucleic acids. Sources of test compounds which can be screened for inhibitory activity against the inventive polypeptides include, for example, libraries of peptides, polypeptides, DNA, RNA and small organic compounds. Additionally, test compounds can be preselected based on a variety of criteria. For example, suitable test compounds for ANSDR1 can be selected as having known short-chain dehydrogenase/reductase inhibition or enhancement activity. Suitable test compounds for TMPRSS2 can be selected as having known serine protease inhibition or enhancement activity. Specific examples of such serine protease inhibitory test compounds include chymostatin, Aprotinin, Propionyl-leupeptin hemisulfate, 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride, and N-(N-Tosyl-L-phenylalanyl)-2-aminoacridone. Alternatively, test compounds can be selected randomly and tested by the screening methods of the present invention. Test compounds are administered to the reaction system at a concentration in the range from about 1 nM to 1 mM. Useful test compounds such as steroids and steroid derivatives are lipophilic, thus allowing them to cross the cell membrane. In addition, routine ligand specific targeting methods are useful for testing compounds for inhibitory activity.

Therefore, the invention provides a method of identifying a compound that inhibits or enhances the activity of an inventive polypeptide where the sample further consists of a prostate cell lysate, a recombinant cell lysate expressing one of the inventive polypeptides, an *in vitro* translation lysate containing mRNA encoding one of the inventive polypeptides, a fractionated sample of a prostate cell lysate, a fractionated sample of a recombinant cell lysate expressing one of the inventive polypeptides, a fractionated sample of an *in vitro* translation lysate containing mRNA encoding one of the inventive polypeptides or an isolated inventive polypeptide. The method can be in single or multiple sample format.

In another embodiment, polypeptides and peptides of the invention can be used as vaccines to prophylactically treat individuals for the occurrence of a prostate neoplastic condition or pathology. Such vaccines can be used to induce B or T cell immune responses or both aspects of the individuals endogenous immune mechanisms. The mode of administration and formulations to induce either or both of these immune responses are well known to those skilled in the art. For example, polypeptides and peptides of the

invention can be administered in many possible formulations, including pharmaceutically acceptable mediums. They can be administered alone or, for example, in the case of a peptide, the peptide can be conjugated to a carrier, such as KLH, in order to increase its immunogenicity. The vaccine can include or be administered in conjunction with an adjuvant, various of which are known to those skilled in the art. After initial immunization with the vaccine, further boosters can be provided if desired. Therefore, the vaccines are administered by conventional methods in dosages which are sufficient to elicit an immunological response, which can be easily determined by those skilled in the art. Alternatively, the vaccines can comprise anti-idiotypic antibodies which are internal images of the inventive polypeptides and peptides described above. Methods of making, selecting and administering such anti-idiotype vaccines are well known in the art. See, for example, Eichmann, et al., *CRC Critical Reviews in Immunology* 7:193-227 (1987).

The invention additionally provides a method of treating or reducing the progression of a prostate neoplastic condition. The method consists of administering to an individual having a neoplastic condition of the prostate an inhibitory amount of an inhibitor specific for a polypeptide of the invention, wherein said inhibitory amount causes a reduction of at least about 2-fold in the amount or activity of the targeted polypeptide. A specific example of a ARSDR1 specific inhibitor is a short-chain dehydrogenase/reductase inhibitor or an ARSDR1 antisense nucleic acid. A specific example of a TMPRSS2 inhibitor is a serine protease inhibitor or a TMPRSS2 antisense nucleic acid. A specific example of an 8C3 specific inhibitor is an 8C3 antisense nucleic acid. Similarly, a specific example of PART-1 specific inhibitor is a PART-1 antisene nucleic acid.

Such inhibitors may be produced using methods which are generally known in the art, and include the use of purified inventive polypeptide to produce antibodies or to screen libraries of compounds, as described previously, for those which specifically bind to one of the inventive polypeptides. For example, known inhibitors of oxidoreductases belonging to the short-chain dehydrogenase/reductase family that inhibit ARSDR1 can be used. Lipophilic compounds able to cross the lipid bilayer that makes up cell membranes are especially useful inhibitors for practicing the methods of the invention.

Antibodies specific to the polypeptides of the present invention can be used, for example, directly as an antagonist, or indirectly as a targeting or delivery mechanism for bringing a cytotoxic or cytostatic agent to neoplastic prostate cells. Such agents can be, for example, radioisotopes. The antibodies can be generated using methods that are well known in the art and include, for example, polyclonal, monoclonal, chimeric, humanized single chain, Fab fragments, and fragments produced by a Fab expression library.

In another embodiment of the invention, the polynucleotides encoding the inventive polypeptides, or any fragment thereof, or antisense molecules, can be used for

therapeutic purposes. In one aspect, antisense molecules to the polynucleotides encoding the polypeptides of the invention can be used to block the transcription or translation of a mRNA homologous to the antisense molecule. Specifically, cells can be transformed with sequences complementary to mRNA transcripts encoding the inventive polypeptides. Such methods are well known in the art, and sense or antisense oligonucleotides or larger polynucleotide fragments, can be designed from various locations along the coding or control regions of sequences encoding the inventive polypeptides. Thus, antisense molecules may be used to modulate the activity of the inventive polypeptides, or to achieve regulation of gene function.

Expression vectors derived from retroviruses, adenovirus, adeno-associated virus (AAV), herpes or vaccinia viruses, or from various bacterial plasmids can be used for delivery of antisense nucleotide sequences to the prostate cell population. The viral vector selected should be able to infect the tumor cells and be safe to the host and cause minimal cell transformation. Retroviral vectors and adenoviruses offer an efficient, useful, and well characterized means of introducing and expressing foreign genes efficiently in mammalian cells. These vectors are well known in the art and have very broad host and cell type ranges, express genes stably and efficiently. Methods which are well known to those skilled in the art can be used to construct such recombinant vectors and are described in Sambrook et al. (*supra*). Even in the absence of integration into the DNA, such vectors can continue to transcribe RNA molecules for a substantial period of time. Transient expression can last for a month or more with a non-replicating vector and even longer if appropriate replication elements are part of the vector system.

Ribozymes, enzymatic RNA molecules, can also be used to catalyze the specific cleavage of mRNAs encoding the polypeptides of the present invention. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to a complementary target RNA, followed by endonucleolytic cleavage. Specific ribozyme cleavage sites within any potential RNA target are identified by scanning the a target RNA for ribozyme cleavage sites which include, for example, the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 10 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site can be evaluated for secondary structural features which can render the oligonucleotide inoperable. The suitability of candidate targets can also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays. Antisense molecules and ribozymes of the invention can be prepared by any method known in the art for the synthesis of nucleic acid molecules.

In another enbodiment, the ARSDR1, TMPRSS2 and PART-1 promoter and regulatory regions can be used for constructing vectors for prostate cancer gene therapy. The promoter and regulatory region can be operably fused to a therapeutic

polynucleotide for prostate specific expression. This method can include the addition of one or more enhancer elements which amplify expression of the heterologous therapeutic polynucleotide without compromising tissue specificity.

Examples of therapeutic polynucleotides that are candidates for prostate gene therapy utilizing the ARSDR1, TMPRSS2 and PART-1 promoters include suicide genes. The expression of suicide genes produces a polypeptide or agent that directly or indirectly inhibits neoplastic prostate cell growth or promotes neoplastic prostate cell death. Suicide genes include genes encoding enzymes such as thymidine kinase, oncogenes, tumor suppressor genes, genes encoding toxins, genes encoding cytokines, or a gene encoding oncostatin. The therapeutic polynucleotides of the present invention can be expressed using the vectors described previously for antisense expression as well as others well known in the art.

It is understood that modifications which do not substantially affect the activity of the various embodiments of this invention are also included within the definition of the invention provided herein. Accordingly, the following examples are intended to illustrate but not limit the present invention.

EXAMPLE 1

Identification of ARSDR1, an Androgen-Regulated Polynucleotide

This example shows identification of ARSDR1, TMPRSS2, PART-1 and 8C3 as genes that are transcriptionally-regulated by androgens in human prostate cancer cells.

To identify genes transcriptionally regulated by androgens, microarrays containing prostate derived cDNAs were screened using RNA from a prostate cell line. Those RNAs showing increased expression levels in response to androgen stimulation were identified and characterized further. Specifically, a non-redundant set of 1500 prostate-derived cDNA clones was identified from the Prostate Expression Database, a public sequence repository of expressed sequence tag (EST) data derived from human prostate cDNA libraries (Hawkins et al., *Nucleic Acids Res.* 27:204-208 (1999)). These 1500 unique cDNAs were sequence verified and the clones were stored in 96 well microtiter plates. The inserts of the cDNAs were amplified with primers BL_ml3F (5'-GTAAAACGACGGCCAGTGAATTG-3') (SEQ ID NO:12) and BL_ml3R (5'-ACACAGGAAACAGCTATGACCATG-3') (SEQ ID NO:13). Two μl of bacteria culture were used as PCR templates. PCR was performed with an initial incubation at 94°C for 5 minutes, followed by 35 cycles of 94°C for 30 seconds, 57°C for 30 seconds, 72°C for 5 minutes, and a final extension at 72°C for 7 minutes. PCR products were purificd with Sephacryl S-500 (Pharmacia, Kalamazoo, MI) on 96-well silent screen filter plates (Nunc, Rochester, NY). The DNA concentration was 200-400 ng/μl. The purified PCR products were mixed with an equal volume of DMSO and spotted twice onto Type IV glass microscope slides (Amersham, Piscataway, NJ) using a

Molecular Dynamics (Sunnyvale, CA) GenII robotic spotting tool. After spotting, the glass slides were air-dried and UV crosslinked with 500 mJ of energy and then baked at 95°C for 30 minutes.

To identify genes transcriptionally regulated by androgens, the microarrays of prostate derived cDNAs were screened using total RNA isolated from LNCap cells cultured for 72 hours either in the presence or absence of a synthetic androgen R1881 (NEN Life Science Products, Boston, MA.

Total RNA was prepared using TRIzol (Gibco-BRL, Germantown, MD) according to the manufacturer's directions. The integrity of the RNA preparation was checked on a standard formaldehyde agarose gel. Fifty μg of the total RNA were digested with 1 μl of RQI RNase-free DNase (Promega, Madison, WI) (1μ/μl) in 1X first strand cDNA synthesis buffer (Gibco-BRL, Germantown, MD) at 37°C for 30 minutes. The reaction mix was then extracted with phenol/chloroform (1:1) and RNA was precipitated with ethanol. The mRNA was isolated from the DNA-free total RNA using a Dynabeads mRNA purification kit (Dynal, Lake Success, NY). LNCaP cells were cultured as follows. The culture medium for LNCap cells was RPMI 1640 with 5% FBS (Gibco-BRL, Germantown, MD). For androgen experiments, 6 flask (175 cm2) of LNCaP cells were starved for androgens by culturing in CS media (RPMI 1640 with 5% of charcoal filtered FBS). After 72 hours of incubation, three flasks were incubated with CS media and the other three were incubated with CS media plus 1 nM of synthetic androgen R1881. All LNCaP cells were incubated for additional 72 hours and then harvested.

Fluorescence-labeled probes were constructed from the above-isolated mRNA as follows. One μg of mRNA or 30 μg of total RNA was mixed with 1μl of anchored oligo dT primer (Amersham, Piscataway, NJ), incubated at 70°C for 10 minutes and then chilled on ice. Four μl of 5X first strand cDNA synthesis buffer (Gibco-BRL, Germantown, MD), 2 μl of 0.1 M DTT (Gibco-BRL, Germantown, MD), 1 μl of HPRI (20 μ/μl) (Amersham, Piscataway, NJ), 1 μl of dNTP mix (Amersham, Piscataway, NJ) containing 2mM dATP, 2mM dGTP, 2mM dTTP and 1mM dCTP, 1 μl of Cy3 dCTP (1mM) (Amersham, Piscataway, NJ) and 1 μl of SuperScript II RT (200 μ/μl) were added. The reactants were incubated at 42°C for 2 hours. Following first strand cDNA labeling, the reaction mixture was incubated at 94°C for 3 minutes. Unlabeled RNA was hydrolyzed by the addition of 1 μl of 5N NaOH and incubation at 37°C for 10 minutes. One μl of 5M HCl and 5 μl of 1M Tris-HCl (pH7.5) were added after the incubation to neutralize the reaction mixture. The mixture was then purified by the Qiagen PCR purification kit (Qiagen, Valencia, CA) following the manufacturer's protocol except washing twice with PE buffer. Following the purification, DNA was eluted with 30 μl of distilled H2O.

Microarray hybridization was performed as follows. One μl of dA/dT (12-18) (1μg/μl) (Pharmacia, Kalamazoo, MI) and 1μl of human Cotl DNA (lμg/μl) (Gibco-BRL, Germantown, MD) were added to the probe. The reaction mixture was then heat denatured at 94°C for 5 minutes. An equal volume of 2X Microarray Hybridization Solution (Amersham, Piscataway, NJ) was added and the mixture was prehybridized at 50°C for 1 hour. After prehybridization, the probe mixture was placed onto a microarray slide with a coverslip. The hybridization was carried out in a humid chamber at 52°C for 16 hours. After hybridization, the slides were washed once with 1X SSC, 0.2% SDS at room temperature for 5 minutes on a shaker, then twice washed with 0.1X SSC, 0.2% SDS at room temperature for 10 minutes. After washing, the slide was dipped into distilled water to remove traces of salt and SDS. Finally, the slides were dried with compressed air.

Analysis of the microarray slides to was performed to identify cDNAs that show increased expression levels in response to androgen stimulation. Hybridized microarray slides were scanned with an Array Scanner Generation II (Amersham, Piscataway, NJ). Intensity data were integrated at a pixel resolution of 10 micrometers using approximately 20 pixels per spot, and recorded at 16 bits. Local background hybridization signals were subtracted prior to comparing spot intensities and determining expression ratios. For each experiment, each cDNA was represented twice on each slide, and the experiments were performed in duplicate producing four data points per cDNA clone per hybridization probe. Intensity ratios for each cDNA clone hybridized with probes derived from androgen-stimulated LNCaP and androgen-starved LNCaP were calculated (stimulated intensity/starved intensity). A gene expression level change was treated as significantly different between the two conditions if all four replicate spots for a given cDNA demonstrated a ratio greater than 2 or less than ½ and the signal intensity was greater than 2 standard deviations above the image background. It had been determined previously that expression ratios less than 2-fold are not reproducible in this system.

Of a total of 1500 distinct cDNAs represented on the microarray 10 cDNAs were identified that upon androgen stimulation exhibited signal intensities at least 1.5 times of local background and exhibited ratios between androgen stimulated and androgen starved cells that were consistently larger than 1.5. These included PSA and hK2, two genes containing androgen response elements located in the 5'-flanking regions that have been shown to confer androgen responsiveness by functional studies (Riegman et al., Molecular Endocrinology 5: 1921-1930 (1991); Murtha et al., Biochemistry 32: 6459-6464 (1993))

Also among the identified cDNAs were four cDNAs that are the subject of the present invention and are referred to as ARSDR1, TMPRSS2, PART-1 and 8C3.

Sequence analysis and BLAST searches of the sequence of cDNA 6A4 against the GenBank databases identified 6A4 as encoding a short chain dehydrogenase/reductase. On this basis the polynucleotide encoded by the 6A4 cDNA was named ARSDR1. The sequence searches also demonstrated that ARSDR1 is a novel polynucleotide. A portion of ARSDR1 matches to Est AA657851 (IMAGE ID: 1207405), but shows no matches to any known genes in the non-redundant subdivision of the GenBank databases.

cDNA 10D11 was found to be homologous to a serine protease termed TMPRSS2. Full-length sequencing of the microarray cDNA confirmed the identity of 10D11 as TMPRSS2 and added additional 3' sequence information to the mRNA sequence available in the public databases. The expression level of TMPRSS2 increased six-fold in androgen stimulated LNCaP cells relative to androgen-deprived cells as assayed by microarray hybridization.

cDNA 14D7, an unknown cDNA as confirmed by BLAST searches against the non-redundant subdivision of the GenBank database. The polynucleotide identified from the initial 14D7 cDNA clone was termed PART-1 for "Prostate Androgen-Regulated Transcript-1".

cDNA 8C3 was also shown to be an unknown cDNA as confirmed by BLAST searches against the non-redundant subdivision of the GenBank database.

EXAMPLE 2

Confirmation of the Androgen-Regulated Expression of ARSDR1

To show up-regulation of ARSDR1 cDNA in response to androgen stimulation a RNA blot containing the same RNAs used for the microarray hybridization was hybridized with ARSDR1 cDNA and control G3PDH cDNA. The RNA blots were made by fractionating 10 μg total RNA on a 1.2% formaldehyde gel and blotting onto nylon filters (Sambrook et al., T. Molecular Cloning, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press (1989)). The ARSDR1 and G3PDH probes were labeled with [α-^{32} P] dCTP (Amersham, Piscataway, NJ) using a rediprime II random primer labeling system (Amersham, Piscataway, NJ) and purified with Sephadex G50 Nick column (Pharmacia, Kalamazoo, MI). RNA hybridization confirmed the microarray hybridization results that ARSDR1 is up-regulated by androgens Quantification by ImageQuant program (Molecular Dynamics, Sunnyvale, CA) revealed that ARSDR1 expression levels in androgen stimulated LNCaP cells is about 15 times higher than in androgen starved LNCaP cells.

To investigate whether the clones obtained as described above represented full length transcripts, 5' rapid amplification of cDNA ends (5'RACE) from human prostate Marathon-Ready cDNA (Clontech, Palo Alto, CA) was performed using primers 6A3_RC3 (5'-GGACAGCATTTTCCTGATTTTGGGGGC-3') (SEQ ID NO:16) and 6A4_RC4(5'-CAGAAGGAGGAGCAACAGCGGGAAC-3') (SEQ ID NO:17).

5' RACE was carried out according to Clontech's protocol. The RACE products were subcloned into PCR2.1-TOPO vectors (Invitrogen, Carlsbad, CA) with the TOPO TA cloning kit (Invitrogen, Carlsbad, CA) and sequenced.

EXAMPLE 3

ARSDR1 is Predominantly Expressed in Prostate Tissue

This example shows the prostate predominant expression and androgen-regulation of ARSDR1.

The expression profile of ARSDR1 in normal human tissues was determined by RNA analysis to determine whether ARSDR1 exhibits tissue specific expression. A multiple tissue Nothern (MTN) blot (Clontech, Palo Alto, CA) containing RNAs from 8 human tissues and an RNA master blot (Clontech, Palo Alto, CA) containing RNAs from 50 human tissues were hybridized with ARSDR1 cDNA probe. The 50 human tissues are: whole brain; amygdala; caudate nucleus; cerebellum; cerebral cortex; frontal lobe, hippocampus; medulla oblongata; occipital lobe; putamen; substantia nigra; temporal lobe; thalamus; acumens; spinal cord; heart; aorta; skeletal muscle; colon; bladder; uterus; prostate; stomach; testis; ovary; pancreas; pituitary gland; adrenal gland; thyroid gland; salivary gland; mammary gland; kidney; liver; small intestine; spleen; thymus; peripheral leukocyte; lymph node; bone marrow; appendix; lung; trachea; placenta; fetal brain; fetal heart; fetal kidney; fetal liver; fetal spleen; fetal thymus; fetal lung; yeast total RNA; yeast tRNA; *E. coli* rRNA; *E. coli* DNA; poly r(A); human Cot1 DNA, human DNA; human DNA; and several no RNA controls. The ARSDR1 cDNA was used as a probe and was labeled with [α-^{32} P] dCTP (Amersham, Piscataway, NJ) using the rediprime II random primer labeling system (Amersham, Piscataway, NJ) followed by purification with Sephadex G50 Nick column (Pharmacia, Kalamazoo, MI). RNA hybridization was carried out in ExpressHyb hybdridization solution (Clontech, Palo Alto, CA). RNA blots were exposed to a phosphor screen (Molecular Dynamics, Sunnyvale, CA) and the images were scanned into a computer with a Phosphorimager. Quantification was done using ImageQuant program (Molecular Dynamics, Sunnyvale).

Overall, not double-counting the six tissues that appeared in both, the MTN blot and the RNA master blot, ARSDR1 expression was analyzed in 52 distinct tissues. Among the 52 tissues analyzed, total ARSDR1 is most abundantly expressed in prostate tissue. It is also slightly expressed in other tissues such as spleen, thymus, testis, ovary, small intestine, colon, peripheral blood leukocyte, and kidney , adrenal gland , and fetal liver.

In addition, a RNA blot containing RNAs from cancer cell lines LNCaP, DU145 and PC3 was made and hybridized with ARSDR1 cDNA probe and G3PDH cDNA control probe. DU145 and PC3 are androgen-unresponsive cell lines. The ARSDR1 cDNA probe was labeled with [α-^{32} P] dCTP (Amersham, Piscataway, NJ) using a

rediprime II random primer labeling system (Amersham, Piscatawaaay, NJ) and purified with Sephadex G50 Nick column (Pharmacia, Kalamazoo, MI). The RNA blot was made by fractionating 10 µg total RNAs on a 1.2% formaldehyde gel and blotting (Sambrook et al., T. Molecular Cloning, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press (1989)). Interestingly, ARSDR1 is expressed both in the androgen-dependent (AD), AR-containing cell line LNCaP and in the androgen-independent, AR-negative cell lines, DU145 and PC3 cells.

EXAMPLE 4

Isolation of the ARSDR1 Full Length cDNA

This example shows the isolation and deduced determination of the nucleotide and deduced amino acid sequence of the ARSDR1 polynucleotide, which contains 2539 base pairs and encodes a 318 aa polypeptide.

To clone a full length ARSDR1 cDNA 1.2 millions phage plaques from a human prostate 5'-stretch cDNA library (Clontech, Palo Alto, CA) were screened with ARSDR1 probe. The screening procedure utilized had been described by Sambook et al. (T. Molecular Cloning, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press (1989)). Five cDNA clones were isolated and sequenced. The inserts of these cDNAs were isolated, subcloned and sequenced. PCR primers 6A4N1 (5'-CCAAAGAGCTGGCTCAGAGAGG-3')(SEQ ID NO:18) and 6A4N2 (5'-CTGGGTGAAGAGGATGTTGGC-3')(SEQ ID NO:19) were designed from the 5' terminus of the existing cDNA, and used to produce a PCR fragment for additional library screening. Eleven additional cDNAs were isolated and sequenced. Furthermore, seven IMAGE cDNA clones (IMAGE CloneID:360400, 109237, 1130518, 1401718, 1337270. 1723130, 1703429) containing ARSDR1 (http://www-bio.llnl.gov/bbrp/image/image.html) were purchased and sequenced.

To investigate whether the clone obtained as described above contained the full length transcript, 5' rapid amplification of cDNA ends (5'RACE) from human prostate Marathon-Ready cDNA (Clontech, Palo Alto, CA) was performed using primers 6A3_RC3 (5'-GGACAGCATTTTCCTGATTTTGGGGGC-3')(SEQ ID NO:16) and 6A4_RC4(5'-CAGAAGGAGGAGCAACAGCGGGAAC-3')(SEQ ID NO:17). 5' RACE was carried out according to Clontech's protocol. The RACE products were subcloned into PCR2.1-TOPO vectors (Invitrogen, Carlsbad, CA) with the TOPO TA cloning kit (Invitrogen, Carlsbad, CA) and sequenced.

Analysis of all of the above clones, revealed a 2539 base pair sequence for ARSDR1 which corresponds to the size of the ARSDR1 transcript as determined by RNA hybridization. ARSDR1 encodes a polypeptide of 318 amino acids (SEQ ID NO:2). The ARSDR1 start codon, has a strong translation start context according to similarity to the Kozak translation initiation consensus sequence (Kozak, *Mamm.*

Genome 7:563-574 (1996)). Two potential polyadenylation signals were identified at nucleotide positions 2439 and 2481. IMAGE clone 1703429 has a poly-A stretch that uses the AATAAA polyadenylation signal at 2419, while ARSDR1 uses the AATAAA signal at 2481. PCR primers flanking the start and stop codons were designed and an expected size band encompassing the coding region was amplified from human prostate Marathon-Ready cDNA (Clontech, Palo Alto, CA).

EXAMPLE 5

ARSDR1 is a Novel Member of the Short-Chain Dehydrogenases/Reductases (SDR)

This example shows that homology searches showed that ARSDR1 is a novel member of the family of short-chain dehydrogenases/reductases (SDR).

BLAST searches were performed and established sequence homology between ARSDR1 and many oxidoreductases from bacteria and plant sources. Subsequently, the polypeptide sequence of ARSDR1 was compared to sequences contained in the BLOCKS database (http://www.blocks.fhcrc.org) (Henikoff and Henikoff, *Nucleic Acids Res.* **19**:6565-6572 (1991)). Blocks are multiply aligned ungapped segments corresponding to the most highly conserved regions of proteins. The BLOCKS database aids in the detection and verification of polypeptide sequence homology by comparing a polypeptide or DNA sequence to a database of polypeptide blocks. The BLOCKS search revealed that the ARSDR1 polypeptide has three blocks that match to short-chain dehydrogenases/reductases (SDR) family protein signature BLOCK (BL00061) with a significant combined E-value of 2.6e-06 (Jornvall et al., *Biochemistry* **34**:6003-6013 (1995)). SDRs are a large family of NAD(H)- or NADP(H)-dependent oxidoreductases, whose members include many enzymes involved in steroid metabolism such as estradiol 17-beta-dehydrogenase (also called 17-beta-hydroxysteroid dehydrogenase) (EC 1.1.1.62), human 15-hydroxyprostaglandin dehydrogenase (NAD+) (EC 1.1.1.141) and 11-beta-hydroxysteroid dehydrogenase (EC 1.1.1.146) (11-DH) (Jornvall et al., *supra* (1995)). Multiple sequence alignments of ARSDR1 with different members of the human hydroxysteroid dehydrogenases (HSD) and a prokaryotic 20 beta-hydroxysteroid dehydrogenase termed Streptomyces 3α/20β-hydroxysteroid dehydrogenase were performed. The alignment was done with the clustalW algorithm (Thompson et al., *Nucl. Acids Res.* **22**:4673-4680 (1994)) from MacVector 6.0 (Oxford Molecular). BLOSUM series matrix, which measures differences between two proteins, was used with an open gap penalty score of 10 and an extend gap penalty score of 0.05. The GenBank accession numbers for the SDR family members used in the alignment are as follows: 20-beta HSD_Strex, Streptomyces 3α/20β-hydroxysteroid dehydrogenase, P19992; 11-beta HSD1_human, P28845; 11-beta HSD2-human, U14631; 17-beta_HSD1_human, P14061; 17-beta_HSSD2-human, L11708; 17-beta_HSD3_human, P37058.

Only two polypeptide motifs were identified as being conserved in the SDR family. The first is a common GlyXXXGlyXGly (SEQ ID NO:14) pattern where the coenzyme NAD(H) or NADP(H) binds at the N terminal of the SDR family enzymes (Jornvall et al., *supra* (1995)). The second motif, TyrXXXLys (SEQ ID NO:15), is indicated to be involved in the catalytic activity of the enzyme (Ghosh et al., *Structure* 2:629-640 (1994)). The ARSDR1 polypeptide contains both of these motifs represented as amino acids 44 to 50 and 198 to 202, respectively in SEQ ID NO:2. Sequence analyses reveal that proteins in the SDR family only exhibit amino acid sequence identity of about 15-30%, likely due to their early divergence and remote origin (Persson et al., *Eur. J. Biochem.* 200:537-543 (1991)). ARSDR1 shows about 25% amino acid sequence identity with other members of the SDR family and was thus determined to be a novel member of the SDR family. Because the polypeptide is androgen regulated and most predominantly expressed in the prostate, it was named androgen regulated short-chain dehydrogenase/reductase 1 (ARSDR1).

Prosite pattern searches revealed that ARSDR1 contains two Asn-glycosylation sites at amino acid positions 174 and 198 (http://www.isrec.isb-sib.ch/software/PSTSCAN_form.html). These two sites are also conserved among SDR family proteins. In addition, two protein kinase C(PKC) phosphorylation sites at amino acid positions 57 and 106, a casein kinase II phosphorylation site at amino acid position 57 and 7 N-myristoylation sites were identified in the ARSDR1 polypeptide.

EXAMPLE 6

Genomic Organization of ARSDR1

This example shows the determination of the ARSDR1 promoter and regulatory as well as coding regions.

To determine the genomic organization of the ARSDR1 polynucleotide ARSDR1 cDNA sequences were aligned against genomic sequences originating from a 197 kb chromosome 14 BAC clone R-1012A1 recently sequenced by the National Sequencing Center-Genoscope in France and deposited to GenBank under accession number AL049779. BAC clone R-1012A1 contains the whole genomic sequence of the ARSDR1 cDNA. The ARSDR1 polynucleotide has 7 exons and 6 introns. The sizes of exons, the sizes of introns, and the exon/intron junctional sequences are listed in Table 1. All the intron/exon junctions conform to the 5'-gt..3'-ag consensus except intron 2, which has a 5'-gc...3'-ag splicing signal (Breathnach and Chambon, *Ann. Rev. Biochem.* 50:349-383 (1981)). The 5'-gc...3'-ag splicing signals have previously been identified in other genes (Devireddy and Jones, *J. Virol.* 72:7294-7301 (1998)).

TABLE 1: Summary of the Genomic Structure of the ARSDR1/6A4 Polynucleotide

Exon	Acceptor	Donor	Exon size (bp)	Intron size (bp)
1		CAGgtctgtgcaatgtattgcca	>114	2577
2	ctctccttctctgtctgcagGAA	GAGgcaagttcacctcctttcaa	120	340
3	tttcatatgttggctgacagGAG	GAGgtaagtgtagaactagagag	159	1193
4	atcgtcttgttccctgcagaGAA	TGGgtaagaaatctggccttatc	104	718
5	attctagtatttctcaacagGTC	AAGgtgggcctagaggaaatgaa	211	5007
6	ttcatgccacccccaaccagGCT	CAGgtatgaatgttatctctttt	191	6591
7	cctttctctttaccttccagTGA		1621	

To characterize the 5' regulatory elements of ARSDR1, 5' genomic sequences were examined for potential transcriptional start sites using a neural network promoter prediction program (http://www-hgc.lbl.gov/projects/promoter.html; Reese et al., Large Scale Sequencing Specific Neural Networks for Promoter and Splice Site Recognition, Biocomputing: Proceedings of the 1996 Pacific Symposium, ed. Lawrence Hunter and Terri E. Klein, World Scientific Publishing Company, Singapore (1996)). A predicted transcription start site 167 base pairs 5' of the ATG start codon was identified. A TATA box (TATAAT) was found 30 base pairs 3' of the putative transcriptional initiation site.

To further characterize the 5' genomic region of ARSDR1 by identifying other potential transcriptional factors the Transcription Element Search Software (TESS) program was utilized (http://www.cbil.upenn.edu/tess/index.html; Schug and Overton, TESS: Transcription Element Search Software on the WWW, Technical Report CBIL-TR-1997-1001-v0.0 of the Computational Biology and Informatics Laboratory, School of Medicine, University of Pennsylvania (1997)). A strong promoter sequence was identified with a score of 0.87. A score of 0.85 has a 0.1-0.4% false positive prediction rate. In addition, a sequence which has 86.7% homology (13 nucleotides out 15) to androgen response element (ARE) consensus sequence (5'-GGA/TACAnnnTGTTCT-3')(SEQ ID NO:20) was identified (Roche et al., *Mol. Endocrinol.* 6:2229-2235 (1992)). Moreover, two sequences which have 86.7% (13 nucleotides out of 15) homology to the consensus sequence of progesterone responsive

elements (PRE) were identified (Lieberman et al., *Mol. Endocrinol.* 7:515-527(1993)). Furthermore, an IL-6 RE-BP (interleukin-6 response element binding protein) site TTCCCAGAA (SEQ ID NO:21) was identified 281 bps 5' of the putative transcription initiation site (Hocke et al., *Mol. Cell Biol.* 12:2282-2294 (1992)).

EXAMPLE 7

Chromosomal Localization of ARSDR1

This example shows the chromosomal localization to human chromosome 14q of ARSDR1 by polymerase chain reaction (PCR).

The medium-resolution Stanford G3 radiation hybrid panel was used to map the chromosome localization of ARSDR1 using primers:

6A4F (5'-GGGGCATTTCCTTACATTGTCCTTG-3') (SEQ ID NO:22) and

6A4R (5'-CACTCCAAACAAGTGATGGGAACAC-3')(SEQ ID NO:23). PCR was performed with 35 cycles of 94°C for 30 seconds, followed by 35 cycles at 54°C for 30 seconds and, finally, 35 cycles at 72°C for 30 seconds. The reaction products were separated on a 1.2% agarose gel and the resulting product pattern was analyzed through the Stanford genome web server (www.shgc.stanford.edu) to determine the probable chromosomal location. ARSDR1 was determined to be localized to SHGC-2558 between two cytogenetically mapped markers D4S63 at 14q23 and D4S258 at 14q24.3 (Genome Database: http://www.gdb.org/). Therefore, ARSDR1 is mapped to 14q23-24.3. This determination is consistent with the fact that the recently sequenced BAC clone R-1012A1 (GenBank accession number: AL049779) containing ARSDR1 comes from chromosome 14q.

EXAMPLE 8

Expression of ARSDR1 in Sections of Normal and Adenocarcinoma Prostate Specimen

This example shows that ARSDR1 is expressed in both normal prostate and prostate carcinoma.

To confirm prostate-specific ARSDR1 expression, *in situ* hybridizations were performed on sections of normal prostate using both, a sense and an antisense RNA probe specific for ARSDR1. A PCR product was generated from the 3' end of the ARSDR1 using primer 6A4insitu1

(5'-TCTTCATTCAGAAAAATTATCTTAG-3')(SEQ ID NO:24) and 6A4insitu2

(5'-GACAGTTCAATATAAATTAAGTAAAAC-3')(SEQ ID NO:25). The PCR product was cloned into PCRII-TOPO (Invitrogen, Carlsbad, CA). The plasmid was then linearized at either end with BamHI or EcoRV, and transcribed to generate sense and anti-sense digoxigenin-labeled probes. Both, dig-dUTP labeled sense and anti-sense probes were constructed using a dig-RNA labeling kit according to manufacturer's instructions (Boehringer Mannheim, Indianapolis, IN). *In situ* hybridization was performed on a Ventana Gen II automated instrument (Ventana Medical Systems,

Tucson, AZ). Formalin-fixed and paraffin-embedded prostate specimens were obtained from a previously surgical specimen tissue bank. The tissue sections (5μm) were mounted onto Proma plus slides (VWR Scientific, W. Chester, PA), deparaffinizeded in a 65°C oven for 2 hours followed by three 5 minutes soaks in xylene and rehydrated through graded alcohol with a final rinse in 2XSSC. Prior to hybridization, the sections were digested with proteinase I cocktail for 12 minutes at 37°C before applying 10 ng of either sense or anti-sense probe in the hybridization buffer. The probe was denatured at 65°C for 4 minutes and hybridized at 42°C for 6 hours. The tissue sections were then rinsed with 2X, 1X and 0.1X SSC at 37°C. The hybridization probe was detected with mouse anti-dig antibody and the signal was amplified by subsequent application of biotin conjugated anti-mouse antibody and streptavidin-horseradish peroxidase. The *in situ* signal was then visualized by DAB and counter-stained with hematoxylin.

ARSDR1 was expressed in both the luminal secretary cells and the basal cells of the epithelia of normal prostate. Little to no hybridization was seen in stromal cells. No background hybridization to normal prostate tissue was seen with the sense ARSDR1 probe.

In situ hybridizations with ARSDR1 sense and antisense probes were also performed on sections of primary prostate adenocarcinoma obtained from radical prostatectomy specimens. ARSDR1 was uniformly expressed in prostate adenocarcinoma cells as revealed by hybridization with anti-sense probes. Hybridization with ARSDR1 sense probes showed no background hybridization to the tumor cells.

EXAMPLE 9

Determination of Androgen-Regulated and Prostate-Localized Expression of TMPRSS2

This example confirms that expression of TMPRSS2 is androgen-regulated and that TMPRSS2 is highly expressed in normal and neoplastic prostate epithelium relative to other human tissues.

TMPRSS2 is a prostate-specific and androgen-regulated polynucleotide that encodes a 492 amino acid serine protease. Androgen-regulated expression of TMPRSS2 was confirmed by Northern analysis using the same LNCaP RNA that was used to construct the probes for microarray hybridization. The LNCaP RNA was isolated using TRIzol (Life Technologies, Germantown, MD) according to the manufacturer's directions. Ten μg of total RNA were fractionated on 1.2% agarose denaturing gels and transferred to nylon membranes by capillary method (Sambrook et al., T. Molecular Cloning, Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press (1989)). Blots were hybridized with TMPRSS2 probes labeled with [alpha-32P]dCTP by random priming using the Random Primers DNA labeling kit (Life Technologies, Germantown, MD) according to the manufacturer's protocol. All DNA manipulations including transformation, plasmid preparation, gel electrophoresis, and probe labeling were

performed according to standard procedures described by Sambrook et al. (T. Molecular Cloning, Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press (1989)). Filters were imaged and quantitated by using a phosphor-capture screen and Imagequant software (Molecular Dynamics, Sunnyvale, CA). Phosphorimage quantitation of the Northern demonstrated a nine-fold induction of TMPRSS2 expression after 72 hours of androgen exposure with synthetic androgen R1881.

TMPRSS2 expression was also studied in the prostate carcinoma cell lines LNCaP, DUI145, and PC3 as well as in androgen-dependent (PXe-AD) and androgen-independent (PXe-AI) prostate cancer xenografts, and prostate stroma (PS). The prostate carcinoma cell lines LNCaP, DUI45, and PC3 were cultured in RPMI 1640 medium supplemented with 10% fetal calf serum (FCS). (Life Technologies, Germantown, MD). Twenty-four hours before androgen regulation experiments, LNCaP cells were transferred into RPMI 1640 media with 10% charcoal-stripped FCS (CS-FCS) (Life Technologies, Germantown, MD). This media was replaced with fresh CS-FCS media or CS-FCS supplemented with 1 nM of the synthetic androgen R1881 (NEN Life Science Products Inc., Boston, MA). Cells were harvested for RNA isolation at 0-, 1-, 2-, 4-, 8-, 24-, 48-, and 72-hour time points. Northern analysis was performed with total RNA isolated from cell lines, normal prostate tissue, and prostate cancer xenografts as described in Example 2.

TMPRSS2 expression could be detected in the normal prostate tissue and the steady-state LNCaP cells grown in FCS, but was not detectable after 24 hours of androgen depletion. Northern blot analysis was performed using a TMPRSS2 probe with RNA extracted from normal prostate (NP), LNCaP at steady state (SS), LNCaP after 24 hours of androgen deprivation (time=0), LNCaP at specified hours after androgen exposure (1, 2, 4, 8, 24, and 48 hours), the PC3 (PC3) and DUI 45 (DU145) prostate cancer cell lines, the androgen dependent (PXe-AD) and androgen-independent (PXe-AI) prostate cancer xenografts, and prostate stroma (PS). TMPRSS2 expression could be detected after 2 hours of androgen supplementation and increased steadily through the 48-hour time point. TMPRSS2 expression was not detectable in the androgen-unresponsive PC-3 and DU-145 cell lines, or in a short-term culture of prostate stroma consisting of fibroblasts and smooth muscle cells.

Normal secretory prostate epithelial cells and early-stage prostate carcinomas depend on androgens for growth. The emergence of an androgen-independent (AI) phenotype is a hallmark of advanced prostate cancer. In addition to AI proliferation, these neoplastic cells are also capable of androgen-independent PSA expression. Northern analysis was further utilized to examine the expression of TMPRSS2 in human prostate cancers propagated in a xenograft system that recapitulates the androgen-dependent (AD) and subsequent AI characteristics of human prostate cancer

growth (Bladou et al., *Int. J. Cancer* 67:785-790 (1996)). TMPRSS2 was expressed in both the AD and AI tumors, a finding that parallels PSA expression in this system, indicating a possible dysregulation of TMPRSS2 control.

The distribution of TMPRSS2 transcripts in normal human tissues was also determined by Northern analysis performed as described in Example 2. Northern blot analysis of TMPRSS2 expression was performed using RNA from 16 human tissues. The human multiple tissue blots were obtained from Clontech (Palo Alto, CA) and contained 2 μg of (poly)A+ RNA in each lane. A beta-actin control probe was used to verify equivalent loading of RNA. Of 16 adult tissues examined, TMPRSS2 message was predominantly expressed in prostate tissues, with very low expression levels in colon, lung, liver, kidney, and pancreas, and no detectable expression in spleen, thymus, testes, ovary, peripheral leukocytes, heart, brain, placenta, or skeletal muscle.

EXAMPLE 10

TMPRSS2 Expression in Prostate Basal Cells and Prostate Carcinoma

This example shows that TMPRSS2 is expressed in prostate basal cells and prostate carcinoma.

Normal prostate contains two major epithelial cell populations, the luminal secretory cells and the basal cells. To localize TMPRSS2 expression, *in situ* hybridizations were performed on sections of normal prostate by using an antisense RNA probe specific for TMPRSS2. For mRNA *in situ* hybridization, recombinant plasmid pCRII-TOPO (Invitrogen, Carlsbad, CA) containing a 489 bp TMPRSS2 fragment (nt 513-1002 of the published TMPRSS2 sequence (Paoloni-Giacobino et al., Genomics 44: 309-20 (1997)) was linearized by restriction digest of the vector to generate sense and antisense digoxigenin-labeled RNA probes. *In situ* hybridization was performed according to the manufacturer's protocol on a Ventana GenII automated instrument (Ventana Medical Systems, Tucson, AZ). Programmed recipe files consisting of buffer rinses, protease digestion, hybridization, detection and counter-stains were optimized for the TMPRSS2 probe. Briefly, the optimized conditions were as follows: Digoxigenin-labeled RNA probe was added manually. Anti-digoxigenin (Ventana Medical Systems, Tucson, AZ) was used as the primary antibody. Denaturation was at 65°C and the hybridization was done at 42°C for 280 minutes. Washes were performed at 35°C with lx, 0.5x, and 0.1x saline sodium citrate (SSC). The system uses a cocktail of anti-rabbit and anti-mouse secondary IgG-biotinylated antibody with an indirect biotin avidin diaminobenzidine (DAB) detection system. The sections were counter-stained with haematoxylin.

The results of the above study showed that TMPRSS2 was expressed exclusively in the normal basal cell population. *In situ* hybridization with an antisense RNA probe for TMPRSS2 was done to assay TMPRSS2 expression in normal and malignant prostate

tissue. TMPRSS2 expression was observed in basal cells of normal prostate tissue, but not in secretory luminal epithelium. The *in situ* images were digitally acquired and the staining intensity was enhanced to show contrast. Little to no staining was seen in stroma, secretory cells, or infiltrating lymphocytes. *In situ* hybridization experiments with sense strand control TMPRSS2 probe showing no background staining in normal prostate tissue. *In situ* hybridizations with TMPRSS2 antisense and sense probes were also performed on sections of primary prostate adenocarcinoma obtained from radical prostatectomy specimens. Adenocarcinoma cells were uniformly positive for TMPRSS2 expression. In addition, TMPRSS2 expression was observed in primary prostate carcinoma cells. The sense strand control TMPRSS2 probe exhibited no background staining in cancerous prostate tissue.

EXAMPLE 11

Sequence Analysis of the Putative TMPRSS2 Promoter

This example shows that the TMPRSS2 polynucleotide contains an androgen response element (ARE) in the 5' promoter region at nucleotides 576 to 590 of SEQ ID NO:9.

To identify androgen regulatory sites, the DNA sequences upstream of the TMPRSS2 coding region was cloned by genome-walking in order. An 1100 base pair DNA fragment overlapping the TMPRSS2 cDNA by 100 nucleotides that contained 870 base pairs of sequence 5' to the putative transcriptional start site was obtained using the GenomeWalker kit by Clontech (Palo Alto, CA). Libraries of adapter-ligated genomic DNA fragments were used as template for PCR reactions with the TMPRSS2 gene-specific primer U75329-71R 5'-TGAGTTCAAAGCCATCTTGCTGTTATCAAC-3' (SEQ ID NO:26) and a primer corresponding to the library adapter sequence AP1 5'-GTAATACGACTCACTATAGGGC-3' (SEQ ID NO:27) according to the manufacturer's instructions. A nested PCR reaction with TMPRSS2 primer U75329-55R 5'-CCATCCTAATACGACTCACTATAGGGC-3' (SEQ ID NO:28) and adapter primer AP2 5'-ACTATAGGGCACGCGTGGT-3' (SEQ ID NO:29) was performed. PCR products were cloned into the pCR2.1 vector (Invitrogen, Carlsbad, CA) and sequenced using Ml3 forward and Ml3 reverse primers. Nucleotide sequences were submitted for homology comparisons against the nonredundant public sequence databases using the BLAST server at the NCBI (http://www.ncbi.nlm.nih.gov/). The BLAST search parameter prompts utilized are the default prompts located at the NCBI BLAST website. Sequences examined for promoter and potential transcriptional start sites using a neural network promoter prediction program (http://www-hgc.lbl.gov/projects/promoter.html) identified a 51 base pair sequence beginning 250 nucleotides 5' of the putative translational start site that correlates highly (score of 0.97 indicating a 0.1 % false-positive prediction rate) with consensus promoter elements. Sequences examined

for transcription factor binding sites using SIGSCAN (http://bimas.dcrt.nih.gov/molbio/signal/) identified numerous putative transcription-factor binding sites including consensus sites for SPl, Z-box, AP1, and AP2 regulation. a 15-bp sequence with significant homology to the consensus androgen response element (ARE) is located at nucleotides 576 to 590 of SEQ ID NO:9.

EXAMPLE 12

Preparation of TMPRSS2-Specific Antibody and Analysis of TMPRSS2 Polypeptide Expression

Polyclonal Antibody

TMPRSS2 peptide sequences were selected by direct primary structure comparison between the members of the serine protease gene family and computer-aided antigenicity, surface probability and hydrophobicity analyses. The TMPRSS2 peptides for antibody production were selected based on the following criteria: 1) the peptide sequence should be on the protein surface and preferably it is in flexible loops; 2) the peptide sequence is at least 15 residues long; and 3) the number of cysteine and proline residues in the selected peptide sequence should kept to a minimum. In order to search for suitable immunogenic peptides in TMPRSS2, the three-dimensional structure of trypsin was evaluated and its loop regions (which are also on the protein surface) were identified. The primary sequence of trypsin was aligned with that of TMPRSS2, and the corresponding loop regions in TMPRSS2 was deduced. Table 2 sets forth the TMPRSS2 peptide sequences that were selected using the these criteria.

TABLE 2. TMPRSS2 Peptides for Immunization

Peptide	Sequence	Residues	Rabbit Sera
SEQ ID NO:30	KVISHPNYDSKTKNNDIC	330-346	6623, 6624
SEQ ID NO:31	KLQKPLTFNDLVKPVC	350-365	6621, 6622
SEQ ID NO:32	CWISGWGATEEKGKTSEV	378-396	6619, 6620

The peptides shown in Table 2 were synthesized and then conjugated with keyhole limpet hemocyanin (KLH) for immunizing rabbits. Conjugated peptides and whole proteins were used for the production of rabbit polyclonal antibodies. These procedures were contracted to the biotechnology company Research Genetics, Inc. (Huntsville, AL). The rabbit anti-TMPRSS2 sera designated in Table 2 were obtained one week following the second boost with each referenced TMPRSS2 peptide antigen. Western blot analysis on lysates from LNCaP cells starved or stimulated with androgens was performed with anti-TMPRSS2 antibody. An induction of TMPRSS2 polypeptide was observed using the anti-TMPRSS2 antibody upon androgen administration. No TMPRSS2 polypeptide was detected in DU145 or PC3 cells which are non-responsive to androgen.

Immunohistochemichal analysis of TMPRSS2 polypeptide expression using polyclonal antibodies raised against the protease domain of the TMPRSS2 polypeptide was performed with normal prostate and prostate carcinoma tissue sections. Normal prostate tissue showed immuno-staining of both basal and luminal epithelial cells. No stromal cell staining was apparent. Prostate carcinoma tissue exhibited variable staining intensity in individual neoplastic cells using TMPRSS2 polyclonal antibody. No reactivity was observed with control non-immune IGG.

Monoclonal Antibody

TMPRSS2 is expressed in mammalian cells in order to produce soluble proteins with suitable post-translational modifications that closely resembles the form of the protein in physiologic sources. The TMPRSS2 full length cDNA sequence shown as SEQ ID NO: 3 is cloned into the plasmid pGT-d (Berg et. al., *Biotechniques* 14:972-978 (1993)) and transfected into the AV12 hamster cell line (ATCC CRL 9595) as described previously for the expression of recombinant hK2 protein (Charlesworth et al., *Urology* 49:487-493 (1997)). Alternatively, the TMPRSS2 cDNAs is cloned into the pLNSX and pLNCX retroviral expression vectors (Miller et al., *Biotechniques* 7:980-2, 984-6, 989-90, (1989)). Stable transfectants are isolated under drug resistance. Individual clones are isolated, expanded, and checked for protein expression by Western blot using the polyclonal antibodies described in Example 12. TMPRSS2 peptides and polypeptide are used to generate monoclonal antibodies by contracting Immgenics Pharmaceuticals, Inc. (Vancouver, British Columbia). Briefly, six-week-old A/J mice (Jackson Laboratories) are immunized with two intraperitoneal injections of selected immunogen, titers are checked and the mice are boosted with intravenous administration of the TMPRSS2 polypeptide. Hybridomas are produced by fusion of mouse splenocytes with P3.653 myeloma cells (Kohler et al, *Nature* 256:495-497 (1975)). Monoclonal antibodies are selected based upon reactivity with TMPRSS2 as well as the failure to react with PSA and hK2 serine proteases using ELISA. Hybridomas are then expanded and antibodies are produced *in vitro* by mass culture or hollow fibers.

ELISA Analysis

An ELISA method for the quantitative screening of patient sera for TMPRSS2 polypeptide is developed using a sandwich ELISA assays as previously described for prostate-specific antigen (PSA) (Corey et al., *Int J Cancer* 71:1019-1028 (1997). Briefly, all combinations of monoclonal or polyclonal antibodies to TMPRSS2 are tested in sandwich assays to determine the pair with the highest sensitivity and specificity. Female sera spiked with different concentrations of recombinant TMPRSS2 polypeptide is used as a control, and to construct a standard curve.

EXAMPLE 13

RNA Blot Analysis of PART-1 Expression

This example corroborates by RNA blot analysis the microarray hybridization results demonstrating androgen-induced up-regulation of PART-1.

A RNA or Northern blot containing the same RNAs used for the microarray hybridization was hybridized to PART-1 cDNA. The RNA blots were made by fractionating 10 μg total RNA on a 1.2% formaldehyde gel and blotting (Sambrook et al., T. Molecular Cloning, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press (1989)). The PART-1 cDNA probe was labeled with [β-^{32} P] dCTP (Amersham, Piscataway, NJ) using a rediprime II random primer labeling system (Amersham, Piscataway, NJ) and the probes were purified with Sephadex G50 Nick column (Pharmacia, Kalamazoo, MI). Northern hybridization confirmed the microarray hybridization results that PART-1 is up-regulated by androgens. The same blot was also hybridized to PSA and G3PDH. PSA was shown to be strongly stimulated by androgens, consistent with previous observation (Montgomery et al., *Prostate* **21**:63-73 (1992)). The amount of RNA loaded on each lane of the northern blot was similar according to G3PDH hybridization. Quantification utilizing the ImageQuant program (Molecular Dynamics, Sunnyvale, CA) revealed that PSA and PART-1 expression levels in androgen stimulated versus androgen starved LNCaP cells are 25.4 and 3.5 times higher, respectively.

The distribution of PART-1 transcripts in normal human tissues was also determined by Northern blot analysis. An RNA Master Blot was purchased from Clontech (Palo Alto, CA) and Northern hybridization was carried out in ExpressHyb hybridization solution (Clontech). The Northern blot was exposed to a phosphor screen (Molecular Dynamics) and the images were scanned into a computer with a Phosphorimager. Quantification was done using ImageQuant program (Molecular Dynamics). Hybridization of PART-1 cDNA probes to a Clonetech RNA Master Blot revealed that PART-1 is expressed most abundantly in prostate with little or no expression detected in colon, lung, liver, kidney, pancreas, spleen, thymus, testes, ovary, peripheral leukocytes, heart, brain, placenta, and skeletal muscle.

EXAMPLE 14

Isolation of the Full Length cDNA for PART-1

This example shows cloning of the full length cDNA for PART-1 and determination of its nucleotide sequence.

To clone PART-1, two rounds of 5' Rapid amplification of cDNA ends (5'RACE) from human Marathon-ready prostate cDNAs (Clontech) and from androgen stimulated LNCaP cDNAs were performed using the Marathon cDNA amplification kit (Clontech) according to manufacturer's protocol. 5' RACE was carried out according to Clontech's protocol. The first round of 5'RACE was performed with primers 14D7-196L (5'-GTGACGGTCTTGGACAGTAAGGG-3')(SEQ ID NO:33) and 14D7-85L

(5'-AGAGTATTGTTGGCTTTGTCTGTC-3')(SEQ ID NO:34). The second round of 5' RACE was performed with primers 14D7RC3 (5-CTTTCCCCTCCGACAAGGAAGCTG-3')(SEQ ID NO:35) and 14D7RC4 (5-CTCATCTGTGTTGTTCCAGTGCAGCC-3')(SEQ ID NO:36). The RACE products were then subcloned into PCR2.1-TOP0 vectors with the TOPO TA cloning kit (Invitrogen) and sequenced. In the second round of RACE using primers 14D7RC3 and 14D7RC4, a 300 bp band was obtained from both, human Marathon-ready prostate cDNAs (Clontech) and androgen stimulated LNCaP cDNAs made by Marathon cDNA amplification kit (Clontech). Sequence analyses of 8 individual RACE clones from both cDNA sources revealed that they all have the same 5' end base, indicating that it is the end of the PART-1 cDNA. Overall, a total of 2109 bp were obtained. This result corresponds to a 2.1 kilobases band that was observed on a Northern blot.

PART-1 cDNA encodes a 60 amino acid polypeptide (SEQ ID NO:6). The translational start site conforms to the Kozak consensus motif for translational start site in an adequate context (Kozak, *Mammalian Genome* 7:563-574 (1996)). The PART-1 polypeptide has no homology to any known proteins in the database by BLAST and FASTA searches. BLOCKS searches (http://www.blocks.fhcrc.org) (Henikoff et al., *Nucleic Acids Research*, 27:204-208 (1999)) revealed that the PART-1 polypeptide has a XPG_1 BLOCK.XPG_1 BLOCK as found in the DNA-damage inducible gene Din7 from yeast (Mieczkowski et al., *Molecular and General Genetics* 253:655-665 (1997)) and in the XPG DNA repair endonuclease (O'Donovan et al., *Journal of Biological Chemistry* 269: 15965-15968(1994)) as well as in exonuclease I of yeast (Fiorentini et al., *Molecular Cell Biology*, 17:2764-2773 (1997)). In addition, the PART-1 polypeptide has two protein kinase C phosphorylation sites and one tyrosine kinase site. Based on BLAST and FASTA database searches it has no homology to any known proteins. A polyadenylation signal AAUAAA (Fitzgerald and Shenk, *Cell* 24:251-260 (1981)) was identified at 633 and 1558 nucleotides 3' of the TAG stop codon. Also, a common natural variant of the polyadenylation signal AUUAAA (Wilusz et al., *Nucleic Acid Research*, 17:3899-3908(1989)) was identified at 644 and 2054 nucleotides 3' of the TAG stop codon (SEQ ID NO:5).

EXAMPLE 15

Isolation of the PART-1 Promoter Region by Genomic Walking

This example shows cloning and sequence analysis of the PART-1 promoter region.

The Human GenomeWalker kit (Clontech) was used to clone the promoter region of the PART-1 cDNA with primers 14D7RC3 and AP1 (5'-GTAATACGACTCACTATAGGGC)(SEQ ID NO:27)(Clontech). Each Genome Walker kit contains five premade "libraries" constructed by digesting human genomic

DNA with 5 enzymes EcoR V, Sca I, Dra I, Pvu II and Ssp I, and ligating the restriction fragments to specific adaptors. PCR was performed with an initial incubation at 94°C for 3 minutes, followed by 5 cycles at 94°C for 25 seconds, followed by 5 cycles at 72°C for 4 minutes, followed by 22 cycles at 94°C for 25 seconds, followed by 22 cycles at 67°C for 4 minutes and a final extension at 67°C for 7 minutes. This genomic walk produced a 1.3 kilobases (kb), a 2.3 kb and 0.8 kb band respectively from the Dra I, Pvu II and Ssp I human GenomeWalking libraries. The 2.3 kb band obtained from the Pvu II library was cloned into a PCR2.1-TOPO vector (Invitrogen) and 2325 base pairs of sequence were obtained. The sequences were examined to identify a potential transcriptional start site using a neural network promoter prediction program (http:www.hgc.lbl.gov/projects/promoter.html).

(Reese et al., Large Scale Sequencing Specific Neural Networks fpr Promoter and Splice Site Recognition, Biocomputing: Proceedings of the 1996 Pacific Symposium, ed. Lawrence Hunter and Terri E. Klein, World Scientific Publishing Company, Singapore (1996)), and for transcriptional factors using the TESS (Transcription Element Search Software) program (http://www.cbil.upenn.edu/tess/index.html) (Schug and Overton, TESS: Transcription Element Search Software on the WWW, Technical Report CBIL-TR-1997-1001-v0.0 of the Computational Biology and Informatics Laboratory, School of Medicine, University of Pennsylvania (1997)). A Dra I and a Ssp I site were found in the sequences corresponding to the respective 1.3 and 0.8 kb genomic walking PCR bands from the Dra I and Ssp I libraries.

The PART-1 genomic walking sequence extends 2024 bps 5' of the start of the PART-1 cDNA. A TATA box (TATAAAA) was identified at nucleotides 1484 to 1491 of SEQ ID NO:11. A putative transcriptional start site (TGTCTTCAAT) is predicted at 30 nucleotides 5' of the TATA box. In addition, a binding site for the homeo-domain containing protein Pbx-1a (Van Dijk et al., *Procl. Nat. Acad. Sci.* (1993)) was identified at nucleotides 536 to 544 of SEQ ID NO:11. The PART-1 promoter region also contains a binding site for NFAT-1 (nuclear factor of activated T cells) at nucleotides 926 to 935 of SEQ ID NO:9 (Rao., *Immunol Today* 15:274-281 (1994)).

Nine putative polymorphisms for the PART-1 polynucleotide were identified (Table 3). These polymorphisms were sequenced verified and either form is represented in at least two clones.

TABLE 3. Summary of the Polymorphisms Found in PART-1

Base Changes

Nucleoetide Position in SEQ ID NO:		Region	Polymorhism
2230	SEQ ID NO:11	Promoter	T -> C

Nucloeotide Position in SEQ ID NO:		Region	Polymorhism
1835	SEQ ID NO:11	Promoter	C -> T
1807	SEQ ID NO:11	Promoter	A -> G
1499	SEQ ID NO:11	Promoter	G -> C
2088	SEQ ID NO:11	Promoter	T -> C
223	SEQ ID NO:5	cDNA	T -> C
589	SEQ ID NO:5	cDNA	T -> C
611	SEQ ID NO:5	cDNA	G -> A
1856	SEQ ID NO:5	cDNA	T -> A

EXAMPLE 16

Chromosomal Localization of PART-1

This example shows the chromosomal localization of PART-1 by both, polymerase chain reaction (PCR) typing and fluorescence *in situ* hybridization (FISH).

The medium-resolution Stanford G3 radiation hybrid panel was used to map the chromosomal localization of PART-1 with primers 14D7mapR (5'-TGCTTTGTTAAGATGAGGCAGGC-3')(SEQ ID NO:37) and 14D7mapF (5'-CATTCCAGGTGTCATGGATAAAGAGC-3')(SEQ ID NO:38). The PCR was performed with an initial incubation at 94°C for 2 minutes, followed by 35 cycles of 94°C for 30 seconds, followed by one cycle at 54°C for 30 seconds and a final cycle at 72°C for 30 seconds. The reaction products were separated on a 1.2% agarose gel and the resulting product pattern was analyzed through the Stanford genome web server (www.shgc.stanford.edu) to determine the probable chromosomal location. Analysis of the typing results indicates that PART-1 is mapped closest to SHGC-14390 on chromosome 5 with a lod score of 8.60 and a cR10,000 distance of 18cRS. SHGC-14390 is mapped between markers D5S2376 and D52604. PART-1 cDNA probe was used to screen an arrayed human BAC genomic library (Research Genetics, Huntsville, AL). Three positive clones 370E12, 493B12 and 508J22 were identified and confirmed by PCR using primers 14D7mapR and 14D7mapF. BAC DNA was biotinylated by nick translation, prehybridized with human Cot 1 DNA (Gibco-BRL) and then hybridized to metaphase spreads of a normal male as described previously (Trask, B.J., Fluorescence *in situ* hybridization , Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1997)). After hybridization and washing, the hybridized sites were labeled with fluorescein-conjugated avidin and the chromosome was counter-stained with DAPI to produce a QFH-like banding pattern. Images were digitized as described elsewhere (Wise et al., Genome Research 7:10-16(1997)). Ten well-spread and well-banded

metaphases were analyzed to localize the hybridization signals. This confirmed that PART-1 is mapped to chromosome 5q12.1.

EXAMPLE 17

Identification, Isolation and Characterization of 8C3, an Androgen-Regulated and Prostate-Specific cDNA

Identification, Characterization, Cloning, and Chromosomal Localization of 8C3 was performed according to essentially the same methods described above in Examples 1, 2 and 5 for ARSDR1.

The chromosomal localization of 8C3 was mapped utilizing primers 8C3mapR (5'-TGGCTTCCTCCCTCCATTTTAGAG-3')(SEQ ID NO:39) and AP1 (Clontech, Palo Alto, CA) in the first round, and primers 8C3mapF (5'-GGTGTCAAAAAACTGGCACATCAG-3')(SEQ ID NO:40) and AP2 (Clontech, Palo Alto, CA) in the second round. The PCR was performed with an initial incubation at 94°C for 30 seconds, followed by one cycle at 54°C for 30 seconds and 35 cycles at 72°C for 30 seconds.

Two cycles of 5' RACE were performed essentially as described above for PART-1 in Example 15 using primer 170L (5'-CTGGAGTGACACAGCGAGACCC-3')(SEQ ID NO:41) in the first round, followed by 5'RACE PCR as follows: one cycle at 94°C for 30 seconds, 5 cycles at 94°C for 5 seconds followed by 72° C for 4 minutes, 5 cycles at 94° C for 5 seconds followed by 70° C for 4 minutes, and 20 cycles at 94°C for 5 seconds followed by 68° C for 4 minutes. In the second round primer 43L (5'-CTGATGTGCCAGTTTTTTGACACC-3')(SEQ ID NO:42) was amplified and 5' RACE PCR was performed as follows: one cycle at 94°C for 30 seconds, 2 cycles at 94°C for 5 seconds followed by 72° C for 4 minutes, and 2 cycles at 94° C for 5 seconds followed by 70° C for 4 minutes.

While the preferred embodiment of the invention has been illustrated and described, it will be appreciated that various changes can be made therein without departing from the spirit and scope of the invention.

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

1. An isolated polynucleotide capable of hybridizing under stringent condition to at least 15 contiguous nucleotides from a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10 and SEQ ID NO:11.

2. A polynucleotide of Claim 1, wherein said fragment comprises substantially the nucleotide sequence shown as nucleotides 1 to 3,113 of SEQ ID NO:8, or functional fragment thereof.

3. A polynucleotide of Claim 2, wherein said functional fragment comprises an androgen response element shown as nucleotide number 2,246 to 2,259 of SEQ ID NO:8.

4. A polynucleotide of Claim 2, wherein said functional fragment comprises a progesterone responsive element shown as nucleotide numbers 2,175 to 2,189 or 2,627 to 2,641 of SEQ ID NO:8.

5. A polynucleotide of Claim 1, wherein said fragment comprises an androgen response element shown as nucleotides 576 to 590 of SEQ ID NO:9.

6. A polynucleotide of Claim 1, wherein said fragment comprises a Pbx-1a regulatory fragment shown as nucleotides 536 to 544 of SEQ ID NO:11.

7. A substantially pure polynucleotide probe comprising at least 15 contiguous nucleotides from a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10 and SEQ ID NO:11.

8. A nucleic acid probe of Claim 7, which comprises an oligonucleotide of 15-18 nucleotides in length.

9. A nucleic acid probe of Claim 7, further comprising a detectable label.

10. A substantially pure polypeptide comprising substantially an amino acid sequence selected from the group consisting of the sequences shown as SEQ ID NO:2, SEQ ID NO:6, and functional fragment thereof.

11. A method of diagnosing or predicting the susceptibility of a prostate neoplastic condition in an individual suspected of having a neoplastic condition of the prostate, comprising:

(a) obtaining a fluid sample from an individual;

(b) determining an expression level of at least one polypeptide chosen from the group consisting of ARSDR1, TMPRSS2, and PART-1; and

(c) comparing said measured expression level of said chosen polypeptide to a normal expression level of said chosen polypeptide from a normal fluid sample, wherein said measured expression level for said chosen polypeptide of 2-fold or more from said fluid sample from said individual compared to said normal expression level indicates the presence of a prostate neoplastic condition.

12. The method of Claim 11, wherein said fluid sample and said normal fluid sample are selected from the group consisting of blood, serum, urine and semen.

13. The method of Claim 11, wherein said expression level is determined by measuring the amount of RNA encoding the chosen polypeptide.

14. The method of Claim 11, wherein said expression level is determined by measuring the activity of said chosen polypeptide.

15. A method of diagnosing or predicting the susceptibility of a prostate neoplastic condition in an individual suspected of having a neoplastic condition of the prostate, comprising:

(a) obtaining a prostate cell sample of the individual;

(b) determining an expression level of at least one polypeptide chosen from ARSDR1, TMPRSS2, and PART-1; and

(c) comparing said measured expression level of said chosen polypeptide to a normal expression level of said chosen polypeptide from normal prostate cells or from an androgen-dependent cell line, wherein said measured expression level for said chosen polypeptide of 2-fold from said individual compared to normal prostate cells or from an androgen-dependent cell line indicates the presence of a prostate neoplastic condition.

16. The method of Claim 15, wherein said expression level is determined by measuring the amount of RNA encoding the chosen polypeptide.

17. The method of Claim 16, wherein said chosen polypeptide is ARSDR1 and the amount of RNA is determined by hybridization with a polynucleotide probe comprising substantially the nucleotide sequence of SEQ ID NO:1, or fragment thereof.

18. The method of Claim 17, wherein said fragment of said polynucleotide probe further comprises an oligonucleotide of about 15-18 nucleotides in length.

19. The method of Claims 18, wherein said polynucleotide probe further comprises a detectable label.

20. The method of Claim 16, wherein said chosen polypeptide is TMPRSS2 and the amount of RNA is determined by hybridization with a polynucleotide probe comprising substantially the nucleotide sequence of SEQ ID NO:3, or fragment thereof.

21. The method of Claim 20, wherein said fragment of said polynucleotide probe further comprises an oligonucleotide of about 15-18 nucleotides in length.

22. The method of Claims 21, wherein said polynucleotide probe further comprises a detectable label.

23. The method of Claim 16, wherein said chosen polypeptide is PART-1 and the amount of RNA is determined by hybridization with a polynucleotide probe comprising substantially the nucleotide sequence of SEQ ID NO:5, or fragment thereof.

24. The method of Claim 23, wherein said fragment of said polynucleotide probe further comprises an oligonucleotide of about 15-18 nucleotides in length.

25. The method of Claims 24, wherein said polynucleotide probe further comprises a detectable label.

26. The method of Claim 15, wherein said chosen polypeptide is ARSDR1 and said amount of polypeptide is determined by contacting a cell, a cell lysate, or fractionated sample thereof, from said individual with a binding agent selective for ARSDR1, and determining the amount of selective binding of said agent.

27. The method of Claim 26, wherein said binding agent selective for ARSDR1 further comprises an antibody or a non-hydrolizable short-chain dehydrogenase/reductase substrate analog.

28. The method of Claim 27, wherein said binding agent further comprises a detectable label.

29. The method of Claim 15, wherein said chosen polypeptide is TMPRSS2 and said amount of polypeptide is determined by contacting a cell, a cell lysate, or

fractionated sample thereof, from said individual with a binding agent selective for TMPRSS2, and determining the amount of selective binding of said agent.

30. The method of Claim 29, wherein said binding agent selective for TMPRSS2 further comprises an antibody or a non-hydrolizable serine protease substrate analog.

31. The method of Claim 30, wherein said binding agent further comprises a detectable label.

32. The method of Claim 15, wherein said chosen polypeptide is PART-1 and said amount of polypeptide is determined by contacting a cell, a cell lysate, or fractionated sample thereof, from said individual with a binding agent selective for PART-1, and determining the amount of selective binding of said agent.

33. The method of Claim 32, wherein said fractionated sample further comprises a lipid membranes.

34. The method of Claim 33, wherein said binding agent selective for PART-1 further comprises an antibody.

35. The method of Claim 34, wherein said binding agent further comprises a detectable label.

36. The method of Claim 15, wherein said expression level is determined by measuring an activity of said chosen polypeptide.

37. The method of Claim 6, wherein said chosen polypeptide is ARSDR1 and said activity is determined by contacting a cell, a cell lysate, or fractionated sample thereof, from said individual with a short-chain dehydrogenase/reductase substrate selective for ARSDR1, and determining the amount of product formed by ARSDR1.

38. The method of Claim 37, wherein the amount of said product formation is determined by measuring the appearance of reduced coenzyme.

39. The method of Claim 37, wherein the amount of said product formation is determined by measuring the disappearance of non-reduced coenzyme.

40. The method of Claim 37, wherein the amount of said product formation is determined by measuring the appearance of said product.

41. The method of Claim 37, wherein the amount of said product formation is determined by measuring the disappearance of said substrate.

42. The method of Claim 36, wherein said chosen polypeptide is TMPRSS2 and said activity is determined by contacting a cell, a cell lysate, or fractionated sample thereof, from said individual with a serine protease substrate selective for TMPRSS2, and determining the amount of product formed by TMPRSS2 .

43. The method of Claim 42, wherein said fractionated sample further comprises lipid membranes.

44. A method of identifying a compound that inhibits the activity of ARSDR1 comprising contacting a sample containing ARSDR1 and a ARSDR1 substrate with a test compound under conditions that allow product formation from said ARSDR1 substrate, and measuring the amount of said product formation from said ARSDR1 substrate, wherein a decrease in the amount of said product formation in the presence of said test compound compared to the absence of said test compound indicates that said compound has ARSDR1 inhibitory activity.

45. The method of Claim 44, wherein the amount of said product formation is determined by measuring the appearance of reduced coenzyme.

46. The method of Claim 44, wherein the amount of said product formation is determined by measuring the disappearance of non-reduced coenzyme.

47. The method of Claim 44, wherein the amount of said product formation is determined by measuring the appearance of said product.

48. The method of Claim 44, wherein the amount of said product formation is determined by measuring the disappearance of said substrate.

49. The method of Claim 44, wherein said sample further comprises prostate tissue, a prostate cell population or a recombinant cell population expressing ARSDR1.

50. The method of Claim 44, wherein said sample further comprises a prostate cell lysate, a recombinant cell lysate expressing ARSDR1, an *in vitro* translation lysate containing ARSDR1 mRNA, a fractionated sample of a prostate cell lysate, a fractionated sample of a recombinant cell lysate expressing ARSDR1, a fractionated sample of an *in vitro* translation lysate containing ARSDR1 mRNA or an isolated ARSDR1 polypeptide.

51. A method of identifying a compound that inhibits the activity of TMPRSS2 comprising contacting a sample containing TMPRSS2 and a TMPRSS2 substrate with a test compound under conditions that allow cleavage of said TMPRSS2 substrate, and measuring the amount of cleavage said TMPRSS2 substrate, a decrease in the amount of cleavage of said TMPRSS2 substrate in the presence of said test compound compared to the absence of said test compound indicates that said compound has TMPRSS2 inhibitory activity.

52. The method of Claim 51, wherein said sample further comprises prostate tissue, a prostate cell population or a recombinant cell population expressing TMPRSS2.

53. The method of claim 51, wherein said sample further comprises a prostate cell lysate, a recombinant cell lysate expressing TMPRSS2, an *in vitro* translation lysate containing TMPRSS2 mRNA, a fractionated sample of a prostate cell lysate, a fractionated sample of a recombinant cell lysate expressing TMPRSS2, a fractionated sample of an *in vitro* translation lysate containing TMPRSS2 mRNA or an isolated TMPRSS2 polypeptide.

54. A method of treating or reducing the progression of a prostate neoplastic condition, comprising administering to an individual having a neoplastic condition of the prostate an inhibitory amount of a selective inhibitor of at least one prostate specific polypeptide chosen from ARSDR1, TMPRSS2, and PART-1, wherein said inhibitory amount causes a reduction of at least about 2-fold in the amount or activity of said chosen polypeptide.

55. The method of Claim 54, wherein said chosen polypeptide is ARSDR1 and said selective inhibitor is a short-chain dehydrogenase/reductase inhibitor.

56. The method of Claim 55, wherein said selective inhibitor is an ARSDR1 antisense polynucleotide.

57. The method of Claim 56, wherein said selective inhibitor binds to the ARSDR1 5' promoter and regulatory region and inhibits transcription of ARSDR1.

58. The method of Claim 54, wherein said chosen polypeptide is TMPRSS2 and said selective inhibitor is a serine protease inhibitor.

59. The method of Claim 58, wherein said selective inhibitor is a TMPRSS2 antisense nucleic acid.

60. The method of Claim 59, wherein said selective inhibitor binds to the TMPRSS2 5' promoter and regulatory region and inhibits transcription of TMPRSS2.

61. The method of Claim 54, wherein said chosen polypeptide is PART-1 and said selective inhibitor is a PART-1 antisense nucleic acid.

62. The method of Claim 61, wherein said selective inhibitor binds to the PART-1 5' promoter and regulatory region and inhibits transcription of PART-1.

63. An antibody that binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, and SEQ ID NO:6, or a fragment thereof.

64. The antibody of Claim 63 wherein the polypeptide has an amino acid sequence substantially similar to the sequence shown in SEQ ID NO:4 and the antibody binds specifically to an epitope in the protease domain of SEQ ID NO:4.

65. The antibody of Claim 64 wherein the epitope is within an amino acid sequence selected from the group consisting of SEQ ID NO:30, SEQ ID NO:31 and SEQ ID NO:32.

SEQUENCE LISTING

<110> University of Washington

<120> Prostate-Specific Polynucleotides, Polypeptides and their Methods of Use

<130> uofw-1-15384

<140>
<141>

<150> 60/130,778
<151> 1999-04-23

<150> 60/151,585
<151> 1999-08-30

<150> 60/174,003
<151> 1999-12-30

<150> 60/177,751
<151> 2000-01-24

<160> 42

<170> PatentIn Ver. 2.0

<210> 1
<211> 2538
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (53)..(994)

<400> 1
gctggagcat cccgctctgg tgccgctgca gccggcagag atggttgagc tc atg ttc 58
Met Phe
1

ccg ctg ttg ctc ctc ctt ctg ccc ttc ctt ctg tat atg gct gcg ccc 106
Pro Leu Leu Leu Leu Leu Leu Pro Phe Leu Leu Tyr Met Ala Ala Pro
5 10 15

```
caa atc agg aaa atg ctg tcc agt ggg gtg tgt aca tca act gtt cag   154
Gln Ile Arg Lys Met Leu Ser Ser Gly Val Cys Thr Ser Thr Val Gln
     20                  25                  30

ctt cct ggg aaa gta gtt gtg gtc aca gga gct aat aca ggt atc ggg   202
Leu Pro Gly Lys Val Val Val Val Thr Gly Ala Asn Thr Gly Ile Gly
 35                  40                  45                  50

aag gag aca gcc aaa gag ctg gct cag aga gga gct cga gta tat tta   250
Lys Glu Thr Ala Lys Glu Leu Ala Gln Arg Gly Ala Arg Val Tyr Leu
                 55                  60                  65

gct tgc cgg gat gtg gaa aag ggg gaa ttg gtg gcc aaa gag atc cag   298
Ala Cys Arg Asp Val Glu Lys Gly Glu Leu Val Ala Lys Glu Ile Gln
             70                  75                  80

acc acg aca ggg aac cag cag gtg ttg gtg cgg aaa ctg gac ctg tct   346
Thr Thr Thr Gly Asn Gln Gln Val Leu Val Arg Lys Leu Asp Leu Ser
         85                  90                  95

gat act aag tct att cga gct ttt gct aag ggc ttc tta gct gag gaa   394
Asp Thr Lys Ser Ile Arg Ala Phe Ala Lys Gly Phe Leu Ala Glu Glu
    100                 105                 110

aag cac ctc cac gtt ttg atc aac aat gca gga gtg atg atg tgt ccg   442
Lys His Leu His Val Leu Ile Asn Asn Ala Gly Val Met Met Cys Pro
115                 120                 125                 130

tac tcg aag aca gca gat ggc ttt gag atg cac ata gga gtc aac cac   490
Tyr Ser Lys Thr Ala Asp Gly Phe Glu Met His Ile Gly Val Asn His
                135                 140                 145

ttg ggt cac ttc ctc cta acc cat ctg ctg cta gag aaa cta aag gaa   538
Leu Gly His Phe Leu Leu Thr His Leu Leu Leu Glu Lys Leu Lys Glu
            150                 155                 160

tca gcc cca tca agg ata gta aat gtg tct tcc ctc gca cat cac ctg   586
Ser Ala Pro Ser Arg Ile Val Asn Val Ser Ser Leu Ala His His Leu
        165                 170                 175

gga agg atc cac ttc cat aac ctg cag ggc gag aaa ttc tac aat gca   634
Gly Arg Ile His Phe His Asn Leu Gln Gly Glu Lys Phe Tyr Asn Ala
    180                 185                 190
```

```
ggc ctg gcc tac tgt cac agc aag cta gcc aac atc ctc ttc acc cag   682
Gly Leu Ala Tyr Cys His Ser Lys Leu Ala Asn Ile Leu Phe Thr Gln
195                 200                 205                 210

gaa ctg gcc cgg aga cta aaa ggc tct ggc gtt acg acg tat tct gta   730
Glu Leu Ala Arg Arg Leu Lys Gly Ser Gly Val Thr Thr Tyr Ser Val
                215                 220                 225

cac cct ggc aca gtc caa tct gaa ctg gtt cgg cac tca tct ttc atg   778
His Pro Gly Thr Val Gln Ser Glu Leu Val Arg His Ser Ser Phe Met
            230                 235                 240

aga tgg atg tgg tgg ctt ttc tcc ttt ttc atc aag act cct cag cag   826
Arg Trp Met Trp Trp Leu Phe Ser Phe Phe Ile Lys Thr Pro Gln Gln
        245                 250                 255

gga gcc cag acc agc ctg cac tgt gcc tta aca gaa ggt ctt gag att   874
Gly Ala Gln Thr Ser Leu His Cys Ala Leu Thr Glu Gly Leu Glu Ile
    260                 265                 270

cta agt ggg aat cat ttc agt gac tgt cat gtg gca tgg gtg tct gtc   922
Leu Ser Gly Asn His Phe Ser Asp Cys His Val Ala Trp Val Ser Val
275                 280                 285                 290

caa gct cgt aat gag act ata gca agg cgg ctg tgg gac gtc agt tgt   970
Gln Ala Arg Asn Glu Thr Ile Ala Arg Arg Leu Trp Asp Val Ser Cys
                295                 300                 305

gac ctg ctg ggc ctc cca ata gac taacaggcag tgccagttgg acccaagaga  1024
Asp Leu Leu Gly Leu Pro Ile Asp
            310

agactgcagc agactacaca gtacttcttg tcaaaatgat tctccttcaa ggttttcaaa 1084

acctttagca caaagagagc aaaaccttcc agccttgcct gcttggtgtc cagttaaaac 1144

tcagtgtact gccagattcg tctaaatgtc tgtcatgtcc agatttactt tgcttctgtt 1204

actgccagag ttactagaga tatcataata ggataagaag accctcatat gacctgcaca 1264

gctcattttc cttctgaaag aaactactac ctaggagaat ctaagctata gcagggatga 1324

tttatgcaaa tttgaactag cttctttgtt cacaattcag ttcctcccaa ccaaccagtc 1384

ttcacttcaa gagggccaca ctgcaacctc agcttaacat gaataacaaa gactggctca 1444
```

ggagcagggc ttgcccaggc atggtggatc accggaggtt agtagttcaa gaccagcctg 1504

gccaacatgg tgaaacccca cctctactaa aaattgttta tatctttgtg tgtctttctg 1564

tttatgtgtg ccaagggagt attttcacaa agttcaaaat agccacaata atcagagatg 1624

gagcaaacca gtgccatcca gtctttatgc aaatgaaatt ctgcaaaggg aagcagattc 1684

tgtatatgtt ggtaactacc caccaagagc acatgggtag cagggaagaa gtaaaaaaag 1744

agaaggagaa tactggaaga taatgcacaa aatgaaggga ctagttaagg attaactagc 1804

cctttaagga ttaactagtt aaggattaat agcaaaagat attaaatatg ctaacatagc 1864

tatggaggaa ttgagggcaa gcacccagga ctgatgaggt cttaacaaaa accagtgtgg 1924

caaaaaaaaa aaaaaaaaaa aaaaaaaaaa atcctaaaaa caaacaaaca aaaaaaacaa 1984

ttcttcattc agaaaaatta tcttagggac tgatattggt aattatggtc aatttaataa 2044

tattttgggg catttcctta cattgtcttg acaagattaa aatgtctgtg ccaaaatttt 2104

gtattttatt tggagacttc ttatcaaaag taatgctgcc aaaggaagtc taaggaatta 2164

gtagtgttcc catcacttgt ttggagtgtg ctattctaaa agattttgat ttcctggaat 2224

gacaattata ttttaacttt ggtgggggaa agagttatag gaccacagtc ttcacttctg 2284

atacttgtaa attaatcttt tattgcactt gttttgacca ttaagctata tgtttagaaa 2344

tggtcatttt acggaaaaat tagaaaaatt ctgataatag tgcagaataa atgaattaat 2404

gttttactta atttatattg aactgtcaat gacaaataaa aattcttttt gattattttt 2464

tgttttcatt taccagaata aaaactaaga attaaaagtt tgattacagt caaaaaaaaa 2524

aaaaaaaaaa aaaa 2538

<210> 2
<211> 314
<212> PRT
<213> Homo sapiens

<400> 2

```
Met Phe Pro Leu Leu Leu Leu Leu Leu Pro Phe Leu Leu Tyr Met Ala
 1               5                  10                  15

Ala Pro Gln Ile Arg Lys Met Leu Ser Ser Gly Val Cys Thr Ser Thr
            20                  25                  30

Val Gln Leu Pro Gly Lys Val Val Val Val Thr Gly Ala Asn Thr Gly
        35                  40                  45

Ile Gly Lys Glu Thr Ala Lys Glu Leu Ala Gln Arg Gly Ala Arg Val
    50                  55                  60

Tyr Leu Ala Cys Arg Asp Val Glu Lys Gly Glu Leu Val Ala Lys Glu
65                  70                  75                  80

Ile Gln Thr Thr Thr Gly Asn Gln Gln Val Leu Val Arg Lys Leu Asp
                85                  90                  95

Leu Ser Asp Thr Lys Ser Ile Arg Ala Phe Ala Lys Gly Phe Leu Ala
            100                 105                 110

Glu Glu Lys His Leu His Val Leu Ile Asn Asn Ala Gly Val Met Met
        115                 120                 125

Cys Pro Tyr Ser Lys Thr Ala Asp Gly Phe Glu Met His Ile Gly Val
    130                 135                 140

Asn His Leu Gly His Phe Leu Leu Thr His Leu Leu Leu Glu Lys Leu
145                 150                 155                 160

Lys Glu Ser Ala Pro Ser Arg Ile Val Asn Val Ser Ser Leu Ala His
                165                 170                 175

His Leu Gly Arg Ile His Phe His Asn Leu Gln Gly Glu Lys Phe Tyr
            180                 185                 190

Asn Ala Gly Leu Ala Tyr Cys His Ser Lys Leu Ala Asn Ile Leu Phe
        195                 200                 205

Thr Gln Glu Leu Ala Arg Arg Leu Lys Gly Ser Gly Val Thr Thr Tyr
    210                 215                 220

Ser Val His Pro Gly Thr Val Gln Ser Glu Leu Val Arg His Ser Ser
225                 230                 235                 240
```

```
Phe Met Arg Trp Met Trp Trp Leu Phe Ser Phe Phe Ile Lys Thr Pro
                245                 250                 255

Gln Gln Gly Ala Gln Thr Ser Leu His Cys Ala Leu Thr Glu Gly Leu
            260                 265                 270

Glu Ile Leu Ser Gly Asn His Phe Ser Asp Cys His Val Ala Trp Val
        275                 280                 285

Ser Val Gln Ala Arg Asn Glu Thr Ile Ala Arg Arg Leu Trp Asp Val
    290                 295                 300

Ser Cys Asp Leu Leu Gly Leu Pro Ile Asp
305                 310


<210> 3
<211> 3966
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (57)..(1535)

<400> 3
gtcatattga acattccaga tacctatcat tactcgatgc tgttgataac agcaag atg 59
                                                              Met
                                                               1

gct ttg aac tca ggg tca cca cca gct att gga cct tac tat gaa aac  107
Ala Leu Asn Ser Gly Ser Pro Pro Ala Ile Gly Pro Tyr Tyr Glu Asn
              5                  10                  15

cat gga tac caa ccg gaa aac ccc tat ccc gca cag ccc act gtg gtc  155
His Gly Tyr Gln Pro Glu Asn Pro Tyr Pro Ala Gln Pro Thr Val Val
         20                  25                  30

ccc act gtc tac gag gtg cat ccg gct cag tac tac ccg tcc ccc gtg  203
Pro Thr Val Tyr Glu Val His Pro Ala Gln Tyr Tyr Pro Ser Pro Val
     35                  40                  45

ccc cag tac gcc ccg agg gtc ctg acg cag gct tcc aac ccc gtc gtc  251
Pro Gln Tyr Ala Pro Arg Val Leu Thr Gln Ala Ser Asn Pro Val Val
```

```
50                  55                  60                  65

tgc acg cag ccc aaa tcc cca tcc ggg aca gtg tgc acc tca aag act   299
Cys Thr Gln Pro Lys Ser Pro Ser Gly Thr Val Cys Thr Ser Lys Thr
                70                  75                  80

aag aaa gca ctg tgc atc acc ttg acc ctg ggg acc ttc ctc gtg gga   347
Lys Lys Ala Leu Cys Ile Thr Leu Thr Leu Gly Thr Phe Leu Val Gly
            85                  90                  95

gct gcg ctg gcc gct ggc cta ctc tgg aag ttc atg ggc agc aag tgc   395
Ala Ala Leu Ala Ala Gly Leu Leu Trp Lys Phe Met Gly Ser Lys Cys
        100                 105                 110

tcc aac tct ggg ata gag tgc gac tcc tca ggt acc tgc atc aac ccc   443
Ser Asn Ser Gly Ile Glu Cys Asp Ser Ser Gly Thr Cys Ile Asn Pro
    115                 120                 125

tct aac tgg tgt gat ggc gtg tca cac tgc ccc ggc ggg gag gac gag   491
Ser Asn Trp Cys Asp Gly Val Ser His Cys Pro Gly Gly Glu Asp Glu
130                 135                 140                 145

aat cgg tgt gtt cgc ctc tac gga cca aac ttc atc ctt cag atg tac   539
Asn Arg Cys Val Arg Leu Tyr Gly Pro Asn Phe Ile Leu Gln Met Tyr
                150                 155                 160

tca tct cag agg aag tcc tgg cac cct gtg tgc caa gac gac tgg aac   587
Ser Ser Gln Arg Lys Ser Trp His Pro Val Cys Gln Asp Asp Trp Asn
            165                 170                 175

gag aac tac ggg cgg gcg gcc tgc agg gac atg ggc tat aag aat aat   635
Glu Asn Tyr Gly Arg Ala Ala Cys Arg Asp Met Gly Tyr Lys Asn Asn
        180                 185                 190

ttt tac tct agc caa gga ata gtg gat gac agc gga tcc acc agc ttt   683
Phe Tyr Ser Ser Gln Gly Ile Val Asp Asp Ser Gly Ser Thr Ser Phe
    195                 200                 205

atg aaa ctg aac aca agt gcc ggc aat gtc gat atc tat aaa aaa ctg   731
Met Lys Leu Asn Thr Ser Ala Gly Asn Val Asp Ile Tyr Lys Lys Leu
210                 215                 220                 225

tac cac agt gat gcc tgt tct tca aaa gca gtg gtt tct tta cgc tgt   779
Tyr His Ser Asp Ala Cys Ser Ser Lys Ala Val Val Ser Leu Arg Cys
                230                 235                 240
```

```
tta gcc tgc ggg gtc aac ttg aac tca agc tgc cag agc agg atc gtg   827
Leu Ala Cys Gly Val Asn Leu Asn Ser Ser Arg Gln Ser Arg Ile Val
            245                 250                 255

ggc ggt gag agc gcg ctc ccg ggg gcc tgg ttc tgg cag gtc agc ctg   875
Gly Gly Glu Ser Ala Leu Pro Gly Ala Trp Phe Trp Gln Val Ser Leu
        260                 265                 270

cac gtc cag aac gtc cac gtg tgc gga ggc tcc atc atc acc ccc gag   923
His Val Gln Asn Val His Val Cys Gly Gly Ser Ile Ile Thr Pro Glu
    275                 280                 285

tgg atc gtg aca gcc gcc cac tgc gtg gaa aaa cct ctt aac aat cca   971
Trp Ile Val Thr Ala Ala His Cys Val Glu Lys Pro Leu Asn Asn Pro
290                 295                 300                 305

tgg cat tgg acg gca ttt gcg ggg att ttg aga caa tct ttc atg ttc  1019
Trp His Trp Thr Ala Phe Ala Gly Ile Leu Arg Gln Ser Phe Met Phe
                310                 315                 320

tat gga gcc gga tac caa gta caa aaa gtg att tct cat cca aat tat  1067
Tyr Gly Ala Gly Tyr Gln Val Gln Lys Val Ile Ser His Pro Asn Tyr
            325                 330                 335

gac tcc aag acc aag aac aat gac att gcg ctg atg aag ctg cag aag  1115
Asp Ser Lys Thr Lys Asn Asn Asp Ile Ala Leu Met Lys Leu Gln Lys
        340                 345                 350

cct ctg act ttc aac gac cta gtg aaa cca gtg tgt ctg ccc aac cca  1163
Pro Leu Thr Phe Asn Asp Leu Val Lys Pro Val Cys Leu Pro Asn Pro
    355                 360                 365

ggc atg atg ctg cag cca gaa cag ctc tgc tgg att tcc ggg tgg ggg  1211
Gly Met Met Leu Gln Pro Glu Gln Leu Cys Trp Ile Ser Gly Trp Gly
370                 375                 380                 385

gcc acc gag gag aaa ggg aag acc tca gaa gtg ctg aac gct gcc aag  1259
Ala Thr Glu Glu Lys Gly Lys Thr Ser Glu Val Leu Asn Ala Ala Lys
                390                 395                 400

gtg ctt ctc att gag aca cag aga tgc aac agc aga tat gtc tat gac  1307
Val Leu Leu Ile Glu Thr Gln Arg Cys Asn Ser Arg Tyr Val Tyr Asp
            405                 410                 415
```

```
aac ctg atc aca cca gcc atg atc tgt ggc ggc ttc ctg cag ggg aac   1355
Asn Leu Ile Thr Pro Ala Met Ile Cys Ala Gly Phe Leu Gln Gly Asn
        420                 425                 430

gtc gat tct tgc cag ggt gac agt gga ggg cct ctg gtc act tcg aac   1403
Val Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Thr Ser Asn
    435                 440                 445

aac aat atc tgg tgg ctg ata ggg gat aca agc tgg ggt tct ggc tgt   1451
Asn Asn Ile Trp Trp Leu Ile Gly Asp Thr Ser Trp Gly Ser Gly Cys
450                 455                 460                 465

gcc aaa gct tac aga cca gga gtg tac ggg aat gtg atg gta ttc acg   1499
Ala Lys Ala Tyr Arg Pro Gly Val Tyr Gly Asn Val Met Val Phe Thr
                470                 475                 480

gac tgg att tat cga caa atg aag gca aac ggc taa tccacatggt        1545
Asp Trp Ile Tyr Arg Gln Met Lys Ala Asn Gly
            485                 490

cttcgtcctt gacgtcgttt tacaagaaaa caatgggggt ggttttgctt ccccgtgcat 1605

gatttactct tagagatgat tcagaggtca cttcattttt attaaacagt gaacttgtct 1665

ggctttggca ctctctgcca tactgtgcag gctgcagtgg ctcccctgcc cagcctgctc 1725

tccctaaccc cttgtccgca aggggtgatg gccggctggt tgtgggcact ggcggtcaat 1785

tgtggaagga agagggttgg aggctgcccc cattgagatc ttcctgctga gtcctttcca 1845

ggggccaatt ttggatgagc atggagctgt cacttctcag ctgctggatg acttgagatg 1905

aaaaaggaga gacatggaaa gggagacagc caggtggcac ctgcagcggc tgccctswsg 1965

ggscacttgg tagtgtcccc agcctacctc tccacaaggg gattttgctg atgggttctt 2025

agagccttag cagccctgga tggtggccag aaataaaggg accagccctt catgggtggt 2085

gacgtggtag tcacttgtaa ggggaacaga aacatttttg ttcttatggg gtgagaatat 2145

agacagtgcc cttggtgcga gggaagcaat tgaaaaggaa cttgccctga gcactcctgg 2205

tgcaggtctc cacctgcaca ttgggtgggg ctcctgggag ggagactcag ccttcctcct 2265

catcctccct gaccctgctc ctagcaccct ggagagtgca catgcccctt ggtcctggca 2325
```

gggcgccaag tctggcacca tgttggcctc ttcaggcctg ctagtcactg gaaattgagg 2385

tccatggggg aaatcaagga tgctcagttt aaggtacact gtttccatgt tatgtttcta 2445

cacattgcta cctcagtgct cctggaaact tagcttttga tgtctcccaag tagtccacct 2505

tcatttaact ctttgaaact gtatcatctt tgccaagtaa gagtggtggc ctatttcagc 2565

tgctttgaca aaatgactgg ctcctgactt aacgttctat aaatgaatgt gctgaagcaa 2625

agtgcccatg gtggcggcga agaagagaaa gatgtgtttt gttttggact ctctgtggtc 2685

ccttccaatg ctgtgggttt ccaaccaggg gaagggtccc ttttgcattg ccaagtgcca 2745

taaccatgag cactactcta ccatggttct gcctcctggc caagcaggct ggtttgcaag 2805

aatgaaatga atgattctac agctaggact taaccttgaa atggaaagtc ttgcaatccc 2865

atttgcagga tccgtctgtg cacatgcctc tgtagagagc agcattccca gggaccttgg 2925

aaacagttgg cactgtaagg tgcttgctcc ccaagacaca tcctaaaagg tgttgtaatg 2985

gtgaaaacgt cttccttctt tattgccccc tcttatttat gtgaacaact gtttgtcttt 3045

ttttgtatct tttttaaact gtaaagttca attgtgaaaa tgaatatcat gcaaataaat 3105

tatgcgattt ttttttcaaa gtaaccactg catctttgaa gttctgcctg gtgagtagga 3165

ccagcctcca tttccttata agggggtgat gttgaggctg ctggtcagag gaccaaaggt 3225

gaggcaaggc cagacttggt gctcctgtgg ttggtgcccc cagttcctgc agcctgtcct 3285

gttggagagg tccctcaaat gactccttct tattattcta ttagtctgtt tccatgctcc 3345

taataaagac atacccaaga ctgcaattta caaaagaaag aagtttattg gatttacaat 3405

tccacatggc tggggtggcc tcacaatcat ggcagaaagc aaggaagagc aaatcacatc 3465

ttacatggat ggcagcaggc agggagagag tttgtgcaca gaaactccca ttttttaaac 3525

catcaggtct tgtgagaccc attcactatc acaaaaacag cacaggaaag acccaccccc 3585

atgattcaat tatctcctat caggtccctc ccacaacaca tgggaattat gggagctaca 3645

agatgagatt tgggtgggga cacagagcca aaccatatca ttctgccccc gcacccccca 3705

aatctcgtat ctttacattt caaaaccaat caggtcttcc caacagtccc ccaaaatctt 3765

aactcatttc agcattaagt caaaagtcta cagtccaaag tctcatccaa cacaaggcaa 3825

gttgcttcca cctatgagcc tgtgaaatca agagcaagtt aattacttcc tagatacaac 3885

gggggtacag gcattgggtg aatacagcca ttccaaatgg gagaaattgg ccaaaataaa 3945

ggggctgcag gccccttaaa a 3966

<210> 4
<211> 492
<212> PRT
<213> Homo sapiens

<400> 4

Met Ala Leu Asn Ser Gly Ser Pro Pro Ala Ile Gly Pro Tyr Tyr Glu
1 5 10 15

Asn His Gly Tyr Gln Pro Glu Asn Pro Tyr Pro Ala Gln Pro Thr Val
20 25 30

Val Pro Thr Val Tyr Glu Val His Pro Ala Gln Tyr Tyr Pro Ser Pro
35 40 45

Val Pro Gln Tyr Ala Pro Arg Val Leu Thr Gln Ala Ser Asn Pro Val
50 55 60

Val Cys Thr Gln Pro Lys Ser Pro Ser Gly Thr Val Cys Thr Ser Lys
65 70 75 80

Thr Lys Lys Ala Leu Cys Ile Thr Leu Thr Leu Gly Thr Phe Leu Val
85 90 95

Gly Ala Ala Leu Ala Ala Gly Leu Leu Trp Lys Phe Met Gly Ser Lys
100 105 110

Cys Ser Asn Ser Gly Ile Glu Cys Asp Ser Ser Gly Thr Cys Ile Asn
115 120 125

Pro Ser Asn Trp Cys Asp Gly Val Ser His Cys Pro Gly Gly Glu Asp
130 135 140

```
Glu Asn Arg Cys Val Arg Leu Tyr Gly Pro Asn Phe Ile Leu Gln Met
145                 150                 155                 160

Tyr Ser Ser Gln Arg Lys Ser Trp His Pro Val Cys Gln Asp Asp Trp
                165                 170                 175

Asn Glu Asn Tyr Gly Arg Ala Ala Cys Arg Asp Met Gly Tyr Lys Asn
            180                 185                 190

Asn Phe Tyr Ser Ser Gln Gly Ile Val Asp Asp Ser Gly Ser Thr Ser
        195                 200                 205

Phe Met Lys Leu Asn Thr Ser Ala Gly Asn Val Asp Ile Tyr Lys Lys
    210                 215                 220

Leu Tyr His Ser Asp Ala Cys Ser Ser Lys Ala Val Val Ser Leu Arg
225                 230                 235                 240

Cys Leu Ala Cys Gly Val Asn Leu Asn Ser Ser Arg Gln Ser Arg Ile
                245                 250                 255

Val Gly Gly Glu Ser Ala Leu Pro Gly Ala Trp Pro Trp Gln Val Ser
            260                 265                 270

Leu His Val Gln Asn Val His Val Cys Gly Gly Ser Ile Ile Thr Pro
        275                 280                 285

Glu Trp Ile Val Thr Ala Ala His Cys Val Glu Lys Pro Leu Asn Asn
    290                 295                 300

Pro Trp His Trp Thr Ala Phe Ala Gly Ile Leu Arg Gln Ser Phe Met
305                 310                 315                 320

Phe Tyr Gly Ala Gly Tyr Gln Val Gln Lys Val Ile Ser His Pro Asn
                325                 330                 335

Tyr Asp Ser Lys Thr Lys Asn Asn Asp Ile Ala Leu Met Lys Leu Gln
            340                 345                 350

Lys Pro Leu Thr Phe Asn Asp Leu Val Lys Pro Val Cys Leu Pro Asn
        355                 360                 365

Pro Gly Met Met Leu Gln Pro Glu Gln Leu Cys Trp Ile Ser Gly Trp
    370                 375                 380
```

```
Gly Ala Thr Glu Glu Lys Gly Lys Thr Ser Glu Val Leu Asn Ala Ala
385                 390                 395                 400

Lys Val Leu Leu Ile Glu Thr Gln Arg Cys Asn Ser Arg Tyr Val Tyr
                405                 410                 415

Asp Asn Leu Ile Thr Pro Ala Met Ile Cys Ala Gly Phe Leu Gln Gly
            420                 425                 430

Asn Val Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Thr Ser
        435                 440                 445

Asn Asn Asn Ile Trp Trp Leu Ile Gly Asp Thr Ser Trp Gly Ser Gly
    450                 455                 460

Cys Ala Lys Ala Tyr Arg Pro Gly Val Tyr Gly Asn Val Met Val Phe
465                 470                 475                 480

Thr Asp Trp Ile Tyr Arg Gln Met Lys Ala Asn Gly
                485                 490



<210> 5
<211> 2108
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (23)..(199)

<400> 5
gattactcac acagtcttga ag atg caa tgt cag cta ttt agg aca gaa aca  52
                       Met Gln Cys Gln Leu Phe Arg Thr Glu Thr
                         1               5                  10

tcc aag gcc gtg tca gaa ctc aat tac gac tac ata tgc att aag gca  100
Ser Lys Ala Val Ser Glu Leu Asn Tyr Asp Tyr Ile Cys Ile Lys Ala
                15                  20                  25

gga act ggc agg cct cag ggt acg cca act ata gga ctc gtg ctt ctc  148
Gly Thr Gly Arg Pro Gln Gly Thr Pro Thr Ile Gly Leu Val Leu Leu
            30                  35                  40
```

gta cgc tgg gct ata atc tat gaa act gag ctc cag agc cag cca atc 196
Val Arg Trp Ala Ile Ile Tyr Glu Thr Glu Leu Gln Ser Gln Pro Ile
45 50 55

act tagctcctca taacaagtct aactggctct ggaaagctga aagggctgca 249
Thr

ctggaacaac acagatgaga tattctacac attaatctac ttatctggaa tcactttgcc 309

tctaaaggcc agagaaaaat cacagcttcc ttgtcggacg ggaaaaggac aggtgatctg 369

gggaaaacgc agctacacct ggagcaaggt ctcttcccgg cttggcaatc tcagctgtgc 429

cggcgctacg ggacccgagc cgtcccagaa accaaagggc aggcacggca gcaaacgcct 489

gagtgctgct gccttcggtg actatatgag aatggaaact tctaaggaag ccaggttgtt 549

agaattgtta ccccctttac tcagagataa catagattat ccaggctgag atggaaaaca 609

agccctttat tgaattttca acacagactc cctgcttctc atctccttaa taaaatttca 669

ttaaaatccc cttgaactcc catgttcaaa tctccatttg ttgacagaca aagccaacaa 729

tactctaaac tgaggcctgc aagtcatttc atttgtattt ttgtccagaa atttcccata 789

ggaagacttc acctcctaca actccgaaga aaacccttac tgtccaagac cgtcaccagc 849

aaccatccgc agtcattcaa gtggaagctt tcacagcttt tgtacattct ctgtgtcaat 909

atacaactga gttacagact gtcccctggc tccctgaccc ttacaaacac taaaagtttt 969

gtttgactca acttcaagct gctcatctgt tagtaagtga tgttcactcc agaacacatt 1029

catgatgaga actttctaaa agaccagcac tgctcttccc ctcctataat cataataatc 1089

atgataacct gaaacatgtt actgggactc gacatttttc tggggattga aatctttagt 1149

ccttggagct gtcacatagc aggggcaacc tcacactgaa acaaaggaag tgatgtccca 1209

ttattatcca ccctgagcca ccataatatg ctgtttacat ttatttttctt cagcctgtgc 1269

aaaacaaagc aatggaaaag gaaactaaaa aatatacata ctagtaccat tatcttcttt 1329

tgcctaaaat tactaatgca ccacgtcagt ctgcttcctt caggcatcat tctcaattca 1389

tcaggacttg tattagcagg ttctggctag agagactatc tcctgtcatc acgatcaatt 1449

aatgttttct ggtgatcaca tcaggcccta tctaagaagc tcatggtata caagggtcac 1509

ccaaatagct gagtgcagtc cttgctcata tttcctttct cttaaccccg caaacaagaa 1569

ttaagatgat cccaataaaa gaaaaattgc tcaggaaact gaacctttct ctgaaccaag 1629

cactgtcagc aaatctcagg tattagagca actatggttg attgaaaagt gtctcaaaat 1689

ctgggccaag aatgattgct aggtccataa gctaatttgt ctggccttgc catttacgta 1749

agccaaagaa agtcactcat gagtaaacta tagaaaacgt tcagacccat cctgttagta 1809

tgtcaaatca actaagactg gcagggtatt aactccattc caggtgtcat ggataaagag 1869

ccccattatt ttcacagtgc cagcctctac ctaaggaaac cctagacctt ggaaccagtt 1929

tcctggtagg gaactgctga cagtttcaat gctgacagtt ggagccaatg cctcatagtg 1989

taaactgaaa gaaaaatagt tgctttttaa aatgtcagca agaaggcctg cctcatctta 2049

acaaagcaaa aaaaaatgct ttaattcaaa ttaaaaatca tgatactaaa aaaaaaaaa 2108

<210> 6
<211> 59
<212> PRT
<213> Homo sapiens

<400> 6
Met Gln Cys Gln Leu Phe Arg Thr Glu Thr Ser Lys Ala Val Ser Glu
1 5 10 15

Leu Asn Tyr Asp Tyr Ile Cys Ile Lys Ala Gly Thr Gly Arg Pro Gln
20 25 30

Gly Thr Pro Thr Ile Gly Leu Val Leu Leu Val Arg Trp Ala Ile Ile
35 40 45

Tyr Glu Thr Glu Leu Gln Ser Gln Pro Ile Thr
50 55

<210> 7

<211> 5217
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (363)
<223> Nucleotide sequence uncertain

<400> 7
ggatcaatat ttttaatttt taatgtattc aattttctgt tccaatcata tttacaacgc 60

tagggtcaat tccgttagct gccgtcattt ctttttgtac tttgattgca ttagccataa 120

actcacgccg caaacgttcc aattcatctt cgttaatctc tgaaaatata ggcgtaccat 180

aatcatcata tccaattgta ttctggacac caccagatgt cataacgata ggtgttgatc 240

cgtcacctgt tagtgtcgtt gatagtcgtg ctaatacatt tccaagtttg tctttgtaat 300

cataaaccaa atttgtttct ttttgtaatt ccatttttтc tccttttatg gtaataataa 360

ctntctactt gacccaccaa acgcaacaac gttatcatga ccaaatgcaa taccaccacc 420

cttattcgtt ccatatgacg ctctaatagt aggcaacttа tctccccccga accagtctcc 480

ccaagatacc cacgtgaacg ccaaatcgtt tttagcgtca ttaatacgca tatcattacc 540

gttgaagtta attctaccgt tgtaaaagtt aaatccagaa cgtgcctcac cagcatacgt 600

tgggttttgc catgtaaaca agtcagtgct gcttatccat gaagccataa ttccaaaccc 660

gattttatct ccaccccaaa tttgaccagc gggagaacca attccgttaa atgtctggtc 720

atataatatt gtcattaatg agttagcagt tgtgttggtt ggactaaatc cagtgaatat 780

tccgccaaca gtttcaaaac cgctactacc atctgcatat tcagtctgtt cgtagaacat 840

cactgctcct gtttcaaatg tagtcgttga cgtaccacgg gtaatcgtca ttgattcagg 900

ttgaatgaac acacttccaa attgtccgtt ccacgctgta ctcataaatg aagaaaagtt 960

accgactaag ctcttaacat taatattaat accattaatt tgagagaagt ctaattgacc 1020

acctttaatg aagttagcgg ttaatgttcc tgtcacgata ttatcagcac ttaaattcat 1080

aacgttaatc ttagcagcgt caatcgtccc gacctgcaac cgagaagcgt cagcagcgcc 1140

tgtaaatttа acccaaaccg tcccagaata agtgtacatt gcttttccgt caggttcgtt 1200

gttgtcattt ttgaaccaaa catcgccaac aactgcgccа actggttcag cgttaccata 1260

aaaaacagta ttgtcacctt gtgcaacagc tctatttttt tgccattgtt gatcagtttt 1320

agcacccaac acccagatag taccatcaaa ataatacttt tctgtattcc cattaggtaa 1380

cgctttgaac catacgtcac ccttttagg gttaggtggt tctactgcac caaatgttgt 1440

attgttgtga ccaccgttag caacactcca ctgtacttga ttgctattag tcgcaatact 1500

tccagataat tgattaatct gtgacgtgat actgttgctt gataagttgt ctcctaatga 1560

caattcagtc atttcaggat ttaacaagtc atacttaact tcaaacactc gtgtcttata 1620

acttaaatca cggtcactat gcataatcaa aactgtgtca ccaagtgaca agttgccgac 1680

cttggctact gttgccgaat actcaactag cgggtgatta atagatacga gtgtttgata 1740

agcttgttga attagcacat tagggtcttc aatttcatca aacgttttca gatataatct 1800

agccgaacca ttaatgtgac catatttagc cgtagcgtca ggatcttcaa gaattttact 1860

tccagcaggt ttatccaatg ggttaccact tgacttagac cacgtaacgc cactgatgtc 1920

aattctacga ccatatccgt caggagagcc gtcagaacct gagctaacct gttccccttt 1980

tccacgtggc aaaatagcgg tgtaaatgct atctgattgc tgtttctttg tgactgttag 2040

caagttactt ccatcagcga aaacttttga cgtgtcgtta ccttgtctga atacatagtc 2100

catgaagcga ccagaaatct tgttgccagt aattgtgata tagaagtaaa cttcgccgcc 2160

cattaaacca acgacttttt ggatagcttc taagtttgta atgtagtaga agtttgttgt 2220

aacactacca ccaatattaa ctgtacctaa cgaccattgc gaaccatcca acgcataagt 2280

catagcttga ctagcattaa cgttgttagg gcgcttgtct tcaatatatg atgttgacgc 2340

aagctcttga taagctaatt cactagctgt gtactcaata atgtcaacgt tgtcagtgcg 2400

tgacataagt cggaaaatca agtgtttatt gctgtcatta gggtgtggtg ttgacacgta 2460

tttcacatca tcattaaacc ctttattaat tggcacactg aatgaaaact cgatagctgt 2520

attaatctta aaatcaatct gaccagccat taaatcgttt tcttttagta gtccgataat 2580

attctgttgt ttatcaaata aataaatcat agaatacgta ccctgtaact aagtgtcaca 2640

tcactgacat tagtacctgt caacttatct ccgttttttg cttttgcttc aaatatattt 2700

gttttttgaa tattaatgtc tgacaacttg cttgttcggt tgattgtgat attgaaattt 2760

accatatcaa tcacaactaa cgcaccagca ggataattac ctaccattgt cagattgtaa 2820

ctgttgttaa tattggcaac aatcgaatgc gttgctgata caagtttcaa agtgattttg 2880

tcgactggtt gtgagtaaat taactgcggg tcattgataa ctgagactgt tgctgagttc 2940

atcaagagga ccatcttgaa aatccccatg aatgaactga caacaatcct gaaggcctgg 3000

gatttttttgt ctgaaaatca actgcagact gtaaatttcc gacagagaaa ggaatctgta 3060

gttcagcact tgatccatct gtgtgaggaa aagcgtgcaa gtatcagtga tgctgccctg 3120

ttagacatca tttatatgca atttcatcag caccagaaag tttgggatgt ttttcagatg 3180

agtaaaggac caggtgaaga tgttgacctt tttgatatga aacaatttaa aaattcgttc 3240

aagaaaattc ttcagagagc attaaaaaat gtgacagtca gcttcagaga aactgaggag 3300

aatgcagtct ggattcgaat tgcctgggga acacagtaca caaagccaaa ccagtacaaa 3360

cctacctacg tggtgtacta ctcccagact ccgtacgcct tcacgtcctc ctccatgctg 3420

aggcgcaata caccgcttct gggtcaggcg ctgacaattg ctagcaaaca ccatcagatt 3480

gtgaaaatgg acctgagaag tcggtatctg gactctctta aggctattgt ttttaaacag 3540

tataatcaga cctttgaaac tcacaactct acgacacctc tacaggaaag aagccttgga 3600

ctagatataa atatggattc aaggatcatt catgaaaaca tagtagaaaa agagagagtc 3660

caacgaataa ctcaagaaac atttggagat tatcctcaac cacaactaga atttgcacaa 3720

tataagcttg aaacgaaatt caaaagtggt ttaaatggga gcatcttggc tgagaggaaa 3780

gaacccctcc gatgcctaat aaagttctct agcccacatc ttctggaagc attgaaatcc 3840

ttagcaccag cgggtattgc agatgctcca ctttttccac tgcttacttg catacccaac 3900

aagagaatga attattttaa aattagagat aaataagacg tgcgtggttt cttaagcaca 3960

gctcctcctt cttgatattg cacatgcact tcagttcatg gctagctgta tagcttccgt 4020

ctgtaaactt gtattttcaa gaatccttgg tattgaattt ttagaaatgc tcacataatt 4080

gttgggactg attcattcct ccacgatatg cctcctctct ctgatatcct gctaactgta 4140

gccgttgtgg catttgagat gacaggacat atatatatat ggcccccacac ttgaccttga 4200

gtgcctgaat gctctgaaat caagcatatg gcacagcgct caagactttt gggtttgtgt 4260

ccttttttct atggctgtct cttctcaatt ctggagaggt ctggttccag tggctggttt 4320

ccagggattg attcttaagc tctggatcac agagagaagc aacaaggaac tatactcaac 4380

tcaaaacttt ttaggagaat catgaaattg gtctattcaa aggatggagt tgagtccatt 4440

ctgttattgt tgcaagaggt tgcatatttg gtgagtcagt tatataaaat agtgttctta 4500

ttgtaaatat gatacttctc ataatctatt ttatcatgtg tataacattc aaactgacaa 4560

atatattgac ttatgaataa aggtgtcaaa aaactggcac atcagttaat tttgatcaaa 4620

gtacttcagt gatcatcact aaatacccta tctttttaaa aatttttcc tttctaattt 4680

tttatttctt tatttattta ttgagacggg gtctcgctgt gtcactccag cctgggtgac 4740

agagtgagac tccgtcttaa aaaataaata aataaaataa aataaatgac atcactttgg 4800

ttcagagctc taaaatggag ggaggaagcc attctaaaaa ggactcccta catgacctgc 4860

aacttgaaaa aaaattaaaa gctccaaaaa aaaaacaata caggagctta ccttgaacct 4920

ttgaattggg ccaaattgcg atgaccactg catcctggaa aatttaattt caccagcact 4980

acaactcctc aacagcacca accaataaac tatggatttt tgtactaagc cagttgcctc 5040

tttcaaaaca acttgtcaac ttgtctaatc accctcagct ttttttaaaa acccctcctc 5100

taccctctct cttcagaaca caagtggctt ctagctgaat ctgtctccca aattgcaatt 5160

cctaagacct caataaaaac accttgtcct tgctaaaaaa aaaaaaaaaa aaaaaaa 5217

<210> 8
<211> 12020
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (2,246)
<223> Androgen response element

<220>
<221> misc_feature
<222> (2,246)
<223> Androgen response element

<220>
<221> misc_feature
<222> (2,175)
<223> Progesterone responsive element

<220>
<221> misc_feature
<222> (2,627)
<223> Progesterone responsive element

<400> 8
aacacatttg ttcaaatatt atctatggct gctttcacat ttacaacagc agagtcgagt 60

agctgcaaca gagaccaaat cctcctcaga gcataaaata tttatttttc ttttcaggag 120

aagtataaca gctcctatac tatccagtcc ctcattattc caatcccatt atagttaata 180

atttttttat gttaaacttt ccctgttcaa attgttgtgt agtttctcac ttctgattga 240

accctgtctg atacatctcc aacctctctg gccgtttctc agtcaatttt tgggctcccc 300

tccctctacc catttcataa tgttagggtt cctcaaagct tccctctgct tttagatagg 360

tttttgctct gccacccagg gtggagtgca gtggcacaat tctagctcac tgcagcctcc 420

aactcctgac tcaaacaatc ctcccacctc agcctcccaa agagctggga ctacagatgt 480

gtaccaccat gcctagctaa ttttttagta gagataatca atggttttta atctcgattc 540

cattatggct ccaatgtgtc tataaaccac ctgaaattat atgcaaattt tatatggata 600

tgcaaataca cacaagtatg gggagaaatt tccagttttc tcttttggca atatagctca 660

atagttaagg gcacaataga ccctgcatac aacttggtgg acaaaagtga ctgcatctgg 720

gatgctaatc tgcaatgctg acttcaatgt tgactgcaat gttgattagc cccagtcctg 780

tgaatatctc ctgattccta ctttatttac tgtcccttgt gtaagaacat gtcaaccttg 840

atgttattgc acaaattatt atctattaca catgtagcat tcttgcctgt tctggaggga 900

tacctttaat tgtcttgcac agagcagttt gactcttttc ctacagtata taaactctga 960

gtgtggagct tacatgtaca gaaatctgcc acccaagccc acatttctgt ctgtaaattc 1020

ccacagtaaa acacccttac caagaaactg gatttgtctg cctcaacctt tggtgtcttg 1080

gctcctctgg catttggggg ccactttgca tatagggccc tctcatggaa caccacctgc 1140

ttcagtttga atcccagctc tgcctctttg acctaggcta gcttacacgg tgcctcagtt 1200

ttcttgtctg taaaataggg aaaattatag tacctacctc ttagagttgt tatgggaatt 1260

aagagttaat atatatatat aaatcattgg gacagggcct ggccataatt aagcagtatt 1320

taagcactgg ctattattgc catttctggc actctccctt ctccctagac ggttttctgt 1380

tgctctagtg caccctgctt tctctgactg tgaaccctca ggtggggtct ttcctctgct 1440

cagcacactc ttgtctatgt tgcttgccta ctaacttccc tctggcctcc gctcaaatgt 1500

tatttcctca aataagcctc tcctgatcta gaactcctct cccttgtaaa agctcccaga 1560

gctttctcta atcctcctca taaccttcac catgatttat ttacagtgat tacttattta 1620

aggtcttctc tgacagtcca tgaactccgg gcgagcagag attatgtgtt cctggcttcc 1680

tgctggacat cagcacagtg cctagtacac tataggaatg ccataaatgt ttgtcaaatg 1740

cgtgaatgac tgaataagga acaaaacaaa tgaaaggctt ttataaagtt ataaaaccca 1800

atacaaatat taggtaacac aataatgaga atgtgacaaa atgtgatgat tgctaggtaa 1860

tctaatgtta aattggtagg aattagtatt gtgtgaagga aaccctctct tctccaattt 1920

gctataaaat atattagttc attggaaccc gtgttcttta accaaatgga aattatgcaa 1980

aaggaccaga aaatattact cggtacaatt aaggaatcac agacagatat tgattggtta 2040

aaacaccctc cactttcacc cctaggaaat aggcctcaga tgaaagaaac tattgaacag 2100

ggaacctatt tattcacaga gaacatattc atggccttaa ttcaagacag gattaaatac 2160

tttcacgagt tggcaaaacc aactgtgcca atctgagata ttcattgcta ccctgggagt 2220

taatgtatgt gacccaagcc atcaggaaga cgatattcta ttatcaatga acactcaaac 2280

actagaaaaa tcatgatagt tgccaccata attcaagata gacacaataa agtgattctc 2340

ctgagaatgt cttcaaggac ttattgcttc ctcttctctg gaaaaaggaa tcctcgagga 2400

ggaagggttt acattccccc accttcccaa ctgtcagact tcggaccgct cctctcctgt 2460

ctaatccgcc tttcatttcc cagacgccaa catctgatct ccaggagaaa cagtgcggta 2520

actgcctaac caggattggg gaaaactgaa ccaggcactg gacttgggtc tggggagata 2580

gaggactgca gagatttcac aaacaccgcg gtgcctagta cactatagga ataagtgtac 2640

tattcctata gtgtatactt ttttttttcc agaaaaatcc ctctggggta tgagcactat 2700

tttcagttct ggttcaactt ttcattgaga ttctatgtgg ataagaggat cttgtagtga 2760

aaaaaagata aaaagcagtg ttttcatctg agtgtgagaa accatgtgtg cctagaggtg 2820

ggagaactga gtattcagaa ggcagaaaaa ataactgaga gctgggtgaa tgatcccaca 2880

ggccctaagc cctagatggt tggggattag cgatgtataa ttcaacttct gccagcgttg 2940

gcggcagcgc tcccacgtag ccagcctgag gctgtcaagc tctggactac atttcccaga 3000

aggtctgctc cccgggctgt cgcaatttga attggggcgt gtctagaaag agaagccata 3060

gtcggcgagc aacgctggag catcccgctc tggtgccgct gcagccggca gagatggttg 3120

agctcatgtt cccgctgttg ctcctccttc tccccttcct tctgtatatg gctgcgcccc 3180

aaatcaggtc tgtgcaaatg tattgccttc tcttgagaat tcttgttgcg gggcgggagg 3240

gaaacataga atccccacca aagctttctt ggaaaactcc cctgggggag ctccaagatg 3300

gcggagccca gagttcggtg ggttggaggc gaggttcagc tgtcctgggt gccaggagga 3360

cacccgagac cccaggcggt gagaaagggc ggcagttccg ggcacactgc ttttccctgg 3420

cctgtttttc ttttgtctaa gcaccccctg attctgggtc cagggaatcc cgagggtaag 3480

gctacctcaa acgcttgagt ttcggactct tcaatgggat atggcccgac gctgttttcc 3540

ccttattccc cacgccttaa accagccttc gaggtacaaa ggagcgattc cacatttcgc 3600

cctctgaaac tgatagcaat cgtaacatct gcggccccat ccttgctctc tcgttgctgt 3660

tacaatcgct tgttctgtgg gcggatgttg taaaggagaa aagtgaaaag gaagagtgtt 3720

tggttatgag ctctatcctg ccttgcgagc aggagcacgg gggatctgag ttaggagcta 3780

gggtactggg gcggtgaggg ctgggagaat ggtgggaagg tattgaccag gtgatagatc 3840

aatcagccca aggctggcca agtgccccag tgaatcagca gcgggctgta aaaccggcag 3900

cctgcttatt ggttctaagt aaatcaccag tggtgacgga ccagtgactg aatagacaga 3960

cacactcctt ccccaggtgg gcggtcttag tcccaataga acttgaagcc aggccagatc 4020

acagtcattt cttaccacag tcacctactt tccttctttt atttttattt tctcttcctt 4080

ctctggtgcc cctcttccag ttatattttt tctgccttct gaacactcag ttctcccacc 4140

tctaggcaca catggtttct cacactcaga tgaaaacact gctttttatc tttttttcac 4200

tacaagatcc tcttatccac atagaatctc aatgaaaagt tgaaccagaa ctgaaaatag 4260

tgctcatacc ccagagggat ttttctggaa aaaaaaaaaa tatatgtgtg tatatatata 4320

tatatatgtg tgtgtgtgta tatatatata tgtgtgtgtg tgtgtatata tatatatata 4380

tatattccag gagctgtatt cccaactcag cagggttatg ggagtcgggc cagagctgga 4440

cttttatagt tgtctctttc attcctgctg gttaaggaga agaaagaact aagtacaggt 4500

cctactaagt tgtcactcct ggagctacat acagtaaggg gatgtgggaa acagcctaga 4560

ggcttgtttt ctgaggaggg aggtgatggt ttgttgggac catgagttaa tctgcctccc 4620

tgtcagcatt ttcgcactat gcaggtagct tcaattggcta agagtatgta ccttgcaatt 4680

ccaagttcaa tgaaaaaggc tgaaatgctc tttaagaatg catctcctgg tcaacagatt 4740

gttttaacca gagtcccaac attgccaaga agacaatagg ccagtctgag aaccaccaat 4800

ctgtgtttgt gcattcattc atttggaaaa ggcaatatag tatagtactt agaaacaagc 4860

actctaggct gggcttggtg actcacacct gtaatcccag cactttggga ggctgaggtg 4920

ggcggatcac ctgaggtcag gagttcgaga ccatcctggc cagcgtggtg aaaccctgtc 4980

tctactaaaa atacaaaaat tagctgggcg tggtggcgca tgggtgaggg ggtgagggat 5040

gagaatcgct tgaacctggg agatgggggt tgcagtggag gtggaggttg cagtgagccg 5100

agatcacacc attgcactcc agcctgggca acaagagcaa aactctgtct caaaaaaaaa 5160

aaaaagagaa agagaaagaa agaaacaagg actctaatta cacaaagcaa gacttgaatc 5220

ccagctcccc tttgtacata cataatcttc tctttgtttt aagtgtgtat gtgcatgcgt 5280

gtgggtgtgt ttttttttte ttgtttattt tgtcccagag tcacaagctt tcataatttt 5340

aacttttgt attggattag ccatttctca aggccataaa atgccagttg tggacagatt 5400

gacctacaga aactagttat ttcttcttct gggctgaatc ttgtgatttc ctaagtcaca 5460

agtgccttag agagctcaaa ccaaagccag gtgcctggct aagcaggtat tcatccagtt 5520

cttgtgctgc agagaacagc tgtaagcttt gcccttgtaa tagaatcact tttgtctttg 5580

agcaaaagca gaaaactttt gtcatctttc ctccaacttt ccttaaatgg acactgtttc 5640

cttaaaagga caacagtgga aagaactaaa gcaaacctca gccagaagac acagggagtt 5700

gcaggaacct actggctgag acagtagaat gaagaattga tgagctctcc ttctctgtct 5760

gcaggaaaat gctgtccagt ggggtgtgta catcaactgt tcagcttcct gggaaagtag 5820

ttgtggtcac aggagctaat acaggtatcg ggaaggagac agcccaaagag ctggctcaga 5880

gaggcaagtt cacctccttt caattccctc ctattacttc tgttttaccc ctcacagtaa 5940

tgatgattca ttctcattgt ttcttcccag ttttggaatt caagaacacc taaggattcc 6000

actgatttgc tgatctgaaa ctagtactat gctctacata taatggaccc ttaatgaata 6060

ttattaattt atttttcttt acctattttgc taatcagata cttctgttgt gaagaaaggg 6120

taaaaatgtg tgttttttaa aaaaatgttg gaaataacat tagatttata atgccagctc 6180

aaaaacagag ctgaccattt ctctttcata tgttggcttg taggagctcg agtatatttа 6240

gcttgccggg atgtggaaaa gggggaattg gtggccaaag agatccagac cacgacaggg 6300

aaccagcagg tgttggtgcg gaaactggac ctgtctgata ctaagtctat tcgagctttt 6360

gctaagggct tcttagctgg taagtgtaga actagagaga ctatgttgtg ggtcttgagg 6420

tcaacatggc agattcaact aaaaggaaaa ggtccctctt cctcatagaa gtggtggata 6480

aaatatagcc aaaaacttct gtagctgaac tcatagtcac tgggacacaa tttatgtcaa 6540

agggatatct tccagaggca gaaacaaaaa atgaattcaa aacgtcatag cagtgagtaa 6600

gagctgaagc cagaaatctc atgggagtct cagcactcag tacaggcttt ttagggtctg 6660

gggtctaaat acttacacag ggataggaga tgaggccaca ggcctttgag agctggagct 6720

gtatttgagg ttctacacaa agttgagatt cacaaagggc catctggtcc atagtataga 6780

agctaataaa aatcctggcc accagctcag agacatgaca aagaagttta ccatctcttt 6840

agacatagat ggatacttac ctgcaaaaaa ttagaacccg agctggcatc atgtgtgggg 6900

tctgaattcg taccaccaaa tgatatgagg atatttatgg gtttgatttt tgtatgtgtg 6960

tgtaatactc tatggaaaat taacgtaaaa acctgtagtg tctcatcaaa agcaaatcca 7020

agagcactct ataggcatat ccccttatag cccagggatc taggcctcca aggaaaaagc 7080

tttcactggt gtgaactcat agtcataaat tacaaaccac atgaggaaat gaaccatgag 7140

agagaattgg aagatgcaat aaagaggtga agccagtatt tatttctttg gggacctgaa 7200

tcccagggca gaggtacata gagaacattc ttttttcttt tgcttttggaa tctttccatt 7260

cagcttggga agtcttttta ttgagagagg tggatgcaac aatttcccaaa agcagtgagt 7320

aagtagctcc tgaactacaa ggatctggaa ggttcatttgg tagtatagaa tggaatttctg 7380

ccccatagta aaagtaggtg ggattcaact aaagtcttaa aatgaaaaac aatagaaaat 7440

aaatatgaag gaaaaatgga gggaagattg aagttctgca tcctcatttt caaagcctga 7500

gcgtgaggca gctttctaag atgtaatagc cttggctagc cacttccacg ctccatcgtc 7560

ttgtccctgc agaggaaaag cacctccacg ttttgatcaa caatgcagga gtgatgatgt 7620

gtccgtactc gaagacagca gatggctttg agatgcacat aggagtcaac cacttgggta 7680

agaaatctgg ccttatcaca aagctagaga gacccaaaat tctctttggg aaacttaggg 7740

gtcagattga ggcttcttgt tgtgtgattg aacctcagga ccataatagg ccatgtgtgc 7800

ataatagttt ggtcctacct gatagctgag gaatacctga gggctgctcc aggttctgcc 7860

cgctttttaa tctgtgattt agatttggtt tgccttggcg catttttaa cagactaatt 7920

ggtttagggg actgtgttgc tttcttagta gtgatccttg tggccagcat caggtatagg 7980

tcatctgctc tttttttgta aagacagggt cttgctatgt tgcccaggct gtacttaaac 8040

tcctggcctc aggcaccttc ctgcctcagc ctctttaagt gctgagatta ccggcgtgag 8100

ccaccgcacc cagccaacca gctcttaaga atgagattaa ccaaagtagg aatgagctgg 8160

gacttcatat tcacatagtc tagacctatc tcgctcatat aaggatgcct catgtctttg 8220

gctgttgtta tttccctaaa ccttcattaa acttgttttc tagttgataa tgaaaagttg 8280

gagggaaaga ccttatttct gaattatctg cctcccatag gtgaggcctg ctcaccacga 8340

ctcctgtcat tcctacatgg ccctgaactt cattaattct agtatttctc aacaggtcac 8400

ttcctcctaa cccatctgct gctagagaaa ctaaaggaat cagccccatc aaggatagta 8460

aatgtgtctt ccctcgcaca tcacctggga aggatccact tccataacct gcagggcgag 8520

aaattctaca atgcaggcct ggcctactgt cacagcaagc tagccaacat cctcttcacc 8580

caggaactgg ccccgagact aaaaggtggg cctagaggaa atgaatgtgg agacatgagt 8640

cagagagagg caggtaaaag ccacaggaat ggaaacttgtt tccagagccc atcacctgga 8700

gaggacctgc ttccgtggct ataattgggt acagttttag taagtagtct acaactgaat 8760

acttggaggtg gctggctggg acagtgagac tctgactttg agtcctgaagg cctggcttta 8820

ctacttacta gctgtgtgat ctatagcaag ttacttaact ttttttttt ttgagacggg 8880

atcttgctct gtcacccatg ctggagtgca atggtgcaat ctcggctcac tgccgacctc 8940

cacttcccag gttcaagtga ttctcctgcc tcagcctcct gagtagctgg gactacaggt 9000

gcgtgccacc aggcctggct aatttttgta tttttagtag agacaggatt tcaccatgtt 9060

ggtcagggtg gtcttgaact cgtgacctca ggtgatccac ccaccttggc ctcccagagt 9120

gctgggattt taggtgtgag ccaccgcacc tggccattac ttaactttta ggacttcctt 9180

cttcatatgt aaatgggact acttcacagt ttattgtcat aattaattca ggcttcataa 9240

gcaagtcaag atttttttact cacagtgaga ctgagtatga gggaaggacc tgaactggca 9300

ctagataaag gcagtagact aaaacaggta ctgggaatat ggaggggcag ggtctccctt 9360

gcaaagcatc aaataagagg gaaatatatg cttgagaagg ggaatttcca gtcacaaata 9420

cacagaggtc cttagcactg tctactctgg atttctaata aagagaaggg aaatcgaggg 9480

tgaactgtgc taggaattga ggcaggcaga atgcaggctc aaggaatcag gcacaggtgc 9540

cctatagttg gaataggatc atcctaccaa atgaacatct gtgagatggg aaggaaactg 9600

gtggttagct cagtccttat gcccagtctg tgcccagggc atgacatctt ttatcagttt 9660

aggatagata agggtaactg gatggtgaaa actatccaaa gagttgatca gggagctagg 9720

gtggcctgga ggtggaaaga tgggtccagt ggtgaagaatt cactaataca tgaatgaggt 9780

gactctgata gttgtgagga ttcagtggaa aatcattatg gtgcagtaaa agaaccctgg 9840

tttgacagaa gattcattca ttcagccatt cagcagacac gtattcagtg cccatgttgt 9900

tccgtgcact gggttagggg acaggggatg aattggtgaa caagataggt atagtctctt 9960

ctttggtttg gtttacatct aacgggggag aaagacaatt agcaagtctt tatacaaact 10020

attaaatgta aaatcagtct agtaatgcat gaataatttt cattgtaaaa agctcaaata 10080

aggtcaggca cagtggctca cacctttaat cccaggcact ttaggaggcc aaaaccggag 10140

gatcacttga ggccagttca agatcagcct gggcaaaata gtgagacctc atctcctcta 10200

ccaaaaaaaa aaaaaaaaga gttcaaagca tatggaagta tgtagaatta aggtgaaaaa 10260

ccaccttcat ttattccccc cttccccctg ctcttctcca ttcatatgac cattatcaac 10320

agcttgtgtg ccccctttcca ataatttttcc tatgcatttt catatgtatg tatcatgcac 10380

aaggaactga aactccaaca cagtatcttg gagatctttc catttagtac atatggattt 10440

acccccttaat tgtattggct gtgttgtatt tcgtagtata gatgtgcctt agtgtattta 10500

tgtatttcct taacaacatc tggattgttt ccttttttgtt gtggctgttt aattataatt 10560

gtgataaatg ttataaagga gaagtaccag gacttacaat ggattttagt gggagtcttg 10620

gggggacctt attcagtctg agggttcagg gaagactttc ctgaagaagt gagttgcatg 10680

ctgagccccg aaggatgatt agaaattaaa tggattagag gtactgatag aatgtggagg 10740

ctgaggccgg gcgcggtggc tcacacctgt aatcccagca ctttgggagg ccgaggctgg 10800

tggatcacct gaggtcagga gttcgagaac atcctggcca acttggtgaa agccctgtct 10860

ctactaaaaa tacaaaaatt agccaggtgt ggtggcgggc gcctgtaatc ctagctactg 10920

gagaggctga ggcaggagaa tcgcctgaac ctgggacaga gtttgcagtg agccaagatc 10980

atgccattgc actccagcct gggcgtcacg agtaaaactc tgtctgaaaa aatgaaaaaa 11040

aaaaagggcc ggacgtggtg gctcatgctt gtaatcccag caatttggga ggctgaggtg 11100

ggtagatcac ctgaggtcag gagttcaaga ccagcctgac caacatggag aaaccccatc 11160

tctactaaaa atacaaaatt agccgggcat gttggcgtgt gcctgtagtc ccagctgctc 11220

aggaggctga ggcaggagaa ttgcttgaac ctgggaggca gaggttgtgg tgagctgaga 11280

tcataccatt gcaccccagc ctgggcaaac aagagtgaaa ctccatctca aaaaaaaaaa 11340

aaaaaaaaag gaatatggag gctgagaaac attccaggca gaaggaggag aatgtacaaa 11400

ggtccgggct cagtttgaga aacagcgaag gttaccattg gtaggtttta gaaagcaaga 11460

gtaataaaga tctgagctga gactggtgag gtaggtagat gccaaaccct tagaggccac 11520

agtaaggatt ttagtgttta ttctaagagc aatgggaaac ggttaaaatt aaagcctttt 11580

caaaatgggg gcaagagggg ttgcagcagt tgatttcatt agcagaggtg agaaatgagg 11640

gtagtttgga ctaggtggtg gaattggagg gtagtagaaa ctttctgagag agatttggtt 11700

tgggaaagct tgttgactgg agatgcatgc tttaggtggg cattcatcag aacccttttg 11760

cctgattggg catgtcggga acactgcaag aaccaggttt ggctttaaaa gtagaaatca 11820

atttgggacc tctgagtctg cattactatg ggtcatctag ggagatgtcg agtaggcagt 11880

tgggaataat acagctgtgt ctgcagctgt cagtccacaa agaccagaaa aattctgttt 11940

cttcaaagct ttacaccaaa actcatttgg cagcaaaaca tttcctgtac tagggagatt 12000

agattaatag ycagtgtatg 12020

<210> 9
<211> 869
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (576)..(590)
<223> Androgen response element

<400> 9
aaaaagtcaa ctgtgagctg tatttggtgc tgagccctga accacccgct gggcagtggg 60

tggctactgg cagtgcctgt ggacgttctc ggggaaagac ggtctgattg cacctgctct 120

tcctcagttc ctagtactgt gaggcactct ttcccactc tcccacaacc ctcagcttaa 180

cacaactgtg gtgattgctt tgtccgtttt cccatgagga ccctgagatt tcaacaaggt 240

tcagtaactc acacatcgac actttccaggg agctgtgttg tcaggaactg gcccaggtct 300

ttccctcctg ggcccatctt gtccttaatc tctgctgacc ttgatgggtt tagttacctt 360

gtcgtgtttc ttcagtgttt tatgccatac agtcttcctt gagtagatgt tctgtatagg 420

ggatatcaca gcttcctggt gagaactggg ttaagtccag caggatgcct gggcttagtg 480

gcactggagt tgtgggtggc cttgagcccc aggcagtccc ctgagagccc ttcccgatgg 540

atgtttgagt cactgagtgt gttccttctt ctttctggtc aggctgtcct ggccgctgca 600

cttacaattg caccccatcc tcctacccac tctgatccca gaggagagtg cttctgctgt 660

aacctctaag gcacaagtgg gtcccaaagt ccttaaatat ggagggatgt ggggaagcag 720

tggtgaagat gcaggccaaa ggtcagcagg caacactggt ctttgttgta ggtaaccttc 780

atctagagat tcttctagtt ctttccagat ttatcttcta aaaactaact ggtatggaaa 840

tattacagtc ctgtaattct ttcttctag 869

<210> 10
<211> 2172
<212> DNA
<213> Homo sapiens

<400> 10
ttgatgtctc caagtagtcc accttcattt aactctttga aactgtatca tctttgccaa 60

gtaagagtgg tggcctattt cagctgcttt gacaaaaatga ctggctcctg acttaacgtt 120

ctataaatga atgtgctgaa gcaaagtgcc catggtggcg gcgaagaaga gaaagatgtg 180

ttttgttttg gactctctgt ggtcccttcc aatgctgtgg gtttccaacc aggggaaggg 240

tccctttttgc attgccaagt gccataacca tgagcactac tctaccatgg ttctgcctcc 300

tggccaagca ggctggtttg caagaatgaa atgaatgatt ctacagctag gacttaacct 360

tgaaatggaa agtcttgcaa tcccatttgc aggatccgtc tgtgcacatg cctctgtaga 420

gagcagcatt cccagggacc ttggaaacag ttggcacttt aaggtgcttg ctccccaaga 480

cacatcctaa aaggtgttgt aatggtgaaa acgttttcct tttttattgc cccttcttat 540

ttatgtgaac aactgtttgt ctttttttgt atttttttta aactgtaaag ttcaattgtg 600

aaaatgaata tcatgcaaat aaattatgcg attttttttt taaagtaacc actgcatctt 660

tgaagttctg cctggtgagt aggaccagcc tccatttcct tataaggggg tgatgttgag 720

gctgctggtc agaggaccaa aggtgaggca aggccagact tggtgctcct gtggttggtg 780

ccctcagttc ctgcagcctg tcctgttgga gaggtccctt aaatgactcc ttcttattat 840

tctattagtc tgtttccatg ctcctaataa agacatacct aagactgcaa tttacaaaag 900

aaagaagttt attggattta caattccaca tggctggggt ggcctcacaa tcatggcaga 960

aagcaaggaa gagcaaatca catcttacat ggatggcagc aggcagggag agagtttgtg 1020

cacagaaact cccatttttt aaaccatcag gtcttgtgag acccattcac tatcacaaaa 1080

acagcacagg aaagacccac ccccatgatt caattatctc ctatcaggtc cctcccacaa 1140

cacatgggaa ttatgggagc tacaagatga gatttgggtg gggacacaga gccaaaccat 1200

atcattctgc ccctgcaccc cccaaatctc gtatctttac atttcaaaac caatcaggtc 1260

ttcccaacag tcccccaaaa tcttaactca tttcagcatt aagtcaaaag tctacagtcc 1320

aaagtctcat ccaacacaag gcaagttgct tccacctatg agcctgtgaa atcaagagca 1380

agttaattac ttcctagata caacgggggt acaggcattg ggtgaataca gccattccaa 1440

atgggagaaa ttggccaaaa taaaggggct gcaggccccc taaaagtcca aaatccagca 1500

gggcagtcaa atcttaaagc accaaaatga cctcctttga ctccatctgt catatccagg 1560

tcacactgat gcaagaggtg ggttcccatg gtctggggca gctccacctc tgtgacttcg 1620

caaggtatag cctccttcct ggctgcatca tgggctggtg ttgagggtct gcagcttttc 1680

caggcacatg gtgcaagctg ttggtggatc taccattctg gggtctggag aatggtggcc 1740

ctcttctcac agctccacta ggcagtgccc cagtagggat tgtgtagggg ctccaacccc 1800

acatttccct tctgcactgc cctagcagag gttttccata agggccccac cctgcagcaa 1860

acttctgcct gggcatctat gcgtttccat acattttctg aaatctaggc agaagttccc 1920

aaatctcaat ccttgacttc tgtgcactgg cagggtcaac atcacatgga agctgccaag 1980

gcttaaggct tgcaacctct gaagccacag ccaaagcttt attttggccc ctttcagcca 2040

tggctggagt ggctaggatg cagggcatca agtttctagg ctgcacacag cacaaggacc 2100

ctgggcctgg cccatgaaac cacttttttct tcctaggcct ctgggcctgt tatgggaggg 2160

gcttctgtga aa 2172

<210> 11
<211> 2385
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (536)..(544)
<223> Pbx-1a regulatory fragment

<400> 11
gcccttgtaa tacgactcac tatagggcac gcgtggtcga cggcccgggc tggtctgatt 60

aatgtttgat aattaagcat ttttatgtga ttcatctcaa taagagccca agcaaggtaa 120

tgtaaacaat gaaggcataa aaaggtggaa ggagatgtga atactacacc aagtgggcca 180

catggaccta aggaaaaaga ggagctgagt tgaagtcagg agctgattca gataaagcag 240

cttcaactct actacttctt actcttagat gggctccaga ccaccttgag ggatgggatt 300

atatttattt tatagatgca ctccttttctc ccactaatac agtcagttag gaatgtatag 360

ctaagaatga gccaaccatt actattcgtg tctatggcta cagaaactca ataaatcaac 420

tcactattct gggtcaataa atgaactcat tatactgttt cttcatgaag taaatgtgtt 480

tctaaaggaa tgtattaaaa ttacattttt aaaaatctat gtattttgca tatatttgat 540

tgatacagaa aaggagtcta ctttggttac ttttcttttt ttttacaaac ttcattcatt 600

ggtgtgatct gaaaggggat aatatttcag aaagcaaaaa tatttttaac aacaaaactt 660

tgtatttcca gtgtgatcta aacacaaaat tatgcaaaac acaactaata tttctctcct 720

actggctaaa tcatgccagt tgtttaggtg caatcaaaaa taagccacta atgttatttt 780

tataggaatt ggcaagtaat tgccaactat agctattaaa agtttatatt tgccacagga 840

gcctgtgatt ttaaatactg ttcaaatgat ttatgcttga gagttataca tagcacaaat 900

acatacattt gtatctacag gtctaaagag gaaaaattag acaaaaagca acttgaaaag 960

gaaccccttt atttttaatg acctaccttt gtgctgtatt gagaaaaagt ataaaaatta 1020

tttgggagga agaaaaatat ttctaagcac taatttcata ggcaaaagcc ctattatttt 1080

ttatgatagg ctaggggaaa tcaataatga tttaaaagca aacagtaaac tgctcacatt 1140

taccagtcta tactcagtat taattctacc ctgcagtgta tgctatagaa ggtaatgaaa 1200

tgctgatgtt ttctggtttt ccacttcaat ctgctccagc tttatcctaa accttgtcag 1260

caacagaagt tacccccag caatcataca aacctaataa cggcttccct ggcttaattt 1320

gatctgctcc aatcacaaat tccaatccac atacttttct ctgttatggt tgtgtctgta 1380

tgtgtttaat gtacttttta taaggctttt ttacaacaag cgcatggttt ttcattcagc 1440

cctctaccac ataaacagaa ccactgaaga gcatttttgt tcatataaaa atgttttagt 1500

tattcatagt gtatgtcttc aattggctgc agattacagt gaacaaaagt attgctgcca 1560

tcattttaca tttctcctgt tactgtgccc tatctgctgt gtacccctgt ccccaaagca 1620

tttattataa tatttgctgc ttgctctcag ttttggactg agatgcttgg gggtgggaaa 1680

acttaaaact tgcctgactc aagtactctc taagctttga tcccagcatc ctgctttaac 1740

ctgctcctgt tggacctctg ccactttact gaccccataa aacaaacaga gctagcaaat 1800

gcatcaaaca gcctccttca aaatgcagct tgaccgtgat tcatggcaca aagctggaat 1860

aaccctgcca gttcttagaa gctttaagat ttatcgtttt gaaattgata cacatagtat 1920

agcaccttaa aaacgaatca tctgctagtt cagtatttct tatttacaat cactccctcc 1980

cagggcatgt aatttatttg ctgagcttaa aaaaaaatat caggattact cacacagtct 2040

tgaagatgca atgttcagct tttaggacag aaacattcta ggccgtgtca gaactcaatt 2100

acgactacat atgcattaag gcaggaactg gcaggccttg gggtacgcca actataggac 2160

tcgtgcttct cgtacgctgg gctataatct atgaaactga gctccagagc cagccaatca 2220

cttagctcct cataacaagt ctaactggct ctggaaagct gaaagggctg cactggaaca 2280

acacagatga gatatttcac acattaatct acttatctgg aatcactttg cctctaaagg 2340

ccagagaaaa atcacagctt ccttgtcgga ggggaaaaga agggc 2385

<210> 12
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<220>
<221> misc_binding
<222> (1)..(23)
<223> PCR primer BL-m13F

<400> 12
gtaaaacgac ggccagtgaa ttg 23

<210> 13
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<220>

<221> misc_binding
<222> (1)..(24)
<223> PCR primer BL-M13R

<400> 13
acacaggaaa cagctatgac catg 24

<210> 14
<211> 7
<212> PRT
<213> Homo sapiens

<220>
<221> BINDING
<222> (1)..(7)
<223> Consensus NAD(H) or NADP(H) binding site domain

<220>
<221> VARIANT
<222> (2)..(4)
<223> Xxx amino acids are not conserved

<220>
<221> VARIANT
<222> (6)
<223> Xxx amino acid is not conserved

<400> 14
Gly Xaa Xaa Xaa Gly Xaa Gly
1 5

<210> 15
<211> 5
<212> PRT
<213> Homo sapiens

<220>
<221> DOMAIN
<222> (1)..(5)
<223> Consensus dehydrogenase/reductase catalytic domain

<220>
<221> VARIANT

<222> (2)..(4)
<223> Xxx amino acids are not conserved

<400> 15
Tyr Xaa Xaa Xaa Lys
1 5

<210> 16
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<220>
<221> misc_binding
<222> (1)..(27)
<223> ARSDR1 RACE forward PCR primer

<400> 16
ggacagcatt ttcctgattt tgggggc 27

<210> 17
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<220>
<221> misc_binding
<222> (1)..(25)
<223> ARSDR1 RACE reverse PCR primer

<400> 17
cagaaggagg agcaacagcg ggaac 25

<210> 18
<211> 22
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding
<222> (1)..(22)
<223> ARSDR1 PCR primer 6A4N1

<400> 18
ccaaagagct ggctcagaga gg 22

<210> 19
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding
<222> (1)..(21)
<223> ARSDR1 PCR primer 6A4N2

<400> 19
ctgggtgaag aggatgttgg c 21

<210> 20
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Androgen
Response Element

<220>
<221> protein_bind
<222> (1)..(15)
<223> Consensus androgen response element

<220>
<221> misc_difference
<222> (7)..(9)
<223> Nucleotide sequence not conserved

<400> 20
ggwacannnt gttct 15

<210> 21
<211> 9
<212> DNA
<213> Homo sapiens

<220>
<221> protein_bind
<222> (1)..(9)
<223> Interleukin response element binding site

<400> 21
ttcccagaa 9

<210> 22
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding
<222> (1)..(25)
<223> ARSDR1 chromosome location primer 6A4F

<400> 22
ggggcatttc cttacattgt ccttg 25

<210> 23
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding
<222> (1)..(25)
<223> ARSDR1 chromosome location primer 6A4R

<400> 23
cactccaaac aagtgatggg aacac 25

<210> 24
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding
<222> (1)..(25)
<223> ARSDR1 PCR primer 6A4insitu1

<400> 24
tcttcattca gaaaaattat cttag 25

<210> 25
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding
<222> (1)..(27)
<223> ARSDR1 PCR Primer 6A4insitu2

<400> 25
gacagttcaa tataaattaa gtaaaac 27

<210> 26
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding
<222> (1)..(30)
<223> TMPRSS2 gene specific primer U-5329-TSR

<400> 26
tgagttcaaa gccatcttgc tgttatcaac 30

<210> 27
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding
<222> (1)..(22)
<223> cDNA library adaptor sequence primer AP1

<400> 27
gtaatacgac tcactatagg gc 22

<210> 28
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding

<222> (1)..(27)
<223> TMPRSS2 gene specific primer U75329-3'R

<400> 28
ccatcctaat acgactcact atagggc 27

<210> 29
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding
<222> (1)..(19)
<223> cDNA library adaptor sequence primer AP2

<400> 29
actatagggc acgcgtggt 19

<210> 30
<211> 18
<212> PRT
<213> Homo sapiens

<220>
<223> Description of Artificial Sequence: PCR Primer

<400> 30
Lys Val Ile Ser His Pro Asn Tyr Asp Ser Lys Thr Lys Asn Asn Asp
1 5 10 15

Ile Cys

<210> 31
<211> 16
<212> PRT
<213> Homo sapiens

<400> 31
Lys Leu Gln Lys Pro Leu Thr Phe Asn Asp Leu Val Lys Pro Val Cys
1 5 10 15

<210> 32
<211> 18
<212> PRT
<213> Homo sapiens

<400> 32
Cys Trp Ile Ser Gly Trp Gly Ala Thr Glu Glu Lys Gly Lys Thr Ser
1 5 10 15

Glu Val

<210> 33
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding
<222> (1)..(23)
<223> PART-1 RACE primer 14D7-196L

<400> 33
gtgacggtct tggacagtaa ggg 23

<210> 34
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding

<222> (1)..(24)
<223> PART-1 RACE primer 14D7-85L

<400> 34
agagtattgt tggctttgtc tgtc 24

<210> 35
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding
<222> (1)..(24)
<223> PART-1 PCR primer 14D7RC3

<400> 35
ctttcccctc cgacaaggaa gctg 24

<210> 36
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding
<222> (1)..(26)
<223> PART-1 PCR primer 14D7RC4

<400> 36
ctcatctgtg ttgttccagt gcagcc 26

<210> 37
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding
<222> (1)..(23)
<223> PART-1 PCR primer 14D7mapR

<400> 37
tgctttgtta agatgaggca ggc 23

<210> 38
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding
<222> (1)..(26)
<223> PART-1 PCR primer 14D7mapF

<400> 38
cattccaggt gtcatggata aagagc 26

<210> 39
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding
<222> (1)..(24)
<223> 8C3 Primer 8C3mapR

<400> 39
tggcttcctc cctccatttt agag 24

<210> 40
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding
<222> (1)..(24)
<223> 8C3 Primer 8C3mapF

<400> 40
ggtgtcaaaa aactggcaca tcag 24

<210> 41
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>
<221> misc_binding
<222> (1)..(22)
<223> 8C3 RACE PCR primer 170L

<400> 41
ctggagtgac acagcgagac cc 22

<210> 42
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR Primer

<220>

<221> misc_binding
<222> (1)..(24)
<223> 8C3 RACE PCR primer 43L

<400> 42
ctgatgtgcc agttttttga cacc 24

THIS PAGE BLANK (USPTO)